EXPERIMENTAL STATISTICS IN ENTOMOLOGY

F. M. WADLEY

GRADUATE SCHOOL PRESS U. S. Department of Agriculture

EXPERIMENTAL STATISTICS IN ENTOMOLOGY

F. M. WADLEY, Ph. D.

Formerly Entomologist and Statistical Consultant, Bureau of Entomology, U. S. Department of Agriculture; now Statistical Consultant to Biological Laboratories, U. S. Department of the Army, and Division Biological Standards, National Institutes of Health.

GRADUATE SCHOOL PRESS
U. S. Department of Agriculture
Washington, D. C.
1967

LIBRARY OF CONGRESS CATALOG NUMBER 67-60083

This book is published and sold by the Graduate School Press, Department of Agriculture, Washington, D. C. 20250.

FOREWORD

The Graduate School takes pride in publishing Dr. Wadley's manuscript on experimental statistics in entomology. This book is a continuation of Dr. Wadley's many contributions to the Graduate School. He has written and taught three correspondence courses in statistics: Statistical Methods in Biology and Agriculture (1939), Experimental Design (1945), and Statistics of Biological Assay (1955). Dr. Wadley is still teaching Experimental Design and Statistics of Biological Assay.

Dr. Wadley has worked as an entomologist and biological statistician since 1914. He has a Ph.D. in entomology from the University of Minnesota. His principal interest and work since 1936 has been in biological statistics. He worked for the Bureau of Entomology, U. S. Department of Agriculture, for 28 years, the Department of the Navy for 4 years, and the Department of the Army for 9 years. He retired from the Federal government in 1962. He now serves as a statistical consultant to Fort Detrick (Army) and the U. S. Public Health Service.

This book is an accumulation of applications developed by Dr. Wadley. A leader in the field of applied statistical analysis, Dr. Wadley presents his material in a comprehensive and concise manner.

John B. Holden
Director
Graduate School

PREFACE

The author conceived the idea of this work between 1940 and 1945, while working as statistical consultant for the Bureau of Entomology. The need of such discussions for those facing statistical questions in experimental work, but not well versed in such methods, was apparent. Since then material has been accumulated as time could be spared from more regular work. The opportunity to draw this work together has come recently, and it is submitted with the hope that it will be helpful.

Many acknowledgments must be made. My teachers, supervisors, and colleagues have made the work possible. I must give credit to the Graduate School for their encouragement; especially to Miss Vera Jensen for much help and counsel, and Mrs. Kathleen Gilbert Franz for careful work in typing. Dr. William Waters, of the Forest Insect Division of the U. S. Forest Service, and Dr. Clifford Maloney of the U. S. Public Health Service, were very helpful in reviewing the manuscript. Mr. William Everard of the U. S. Forest Service gave valuable editorial help.

F. M. Wadley

CONTENTS

INTRODUCTION

This discussion is intended to help entomological workers who have a real interest, but little training, in applying statistical methods to research problems. As such it is very elementary; mathematicians and entomologists advanced in statistical methods will probably be dissatisfied with it. However, it is the author's belief that this level of presentation will do the most good. It is a general observation that when research workers begin to analyze their own data, improvement in experiments and results soon follows. After trying out analytical methods, references to more advanced methods can be utilized, and this discussion will have served its purpose.

In our text, a survey of problems, and of simple computation methods, is followed by a chapter on sampling principles and one on experimental design. Last, a collection of practical problems that have actually arisen is recorded, with analytical procedures shown in detail.

In applying statistics to science, we should remember the classification of sciences into physical, biological, and social, in decreasing order of exactness. Our biological material shows more variation than that of the physicist or chemist. Their standards of exactness are usually not practicable in biology.

Statistical analysis has been developed, from considerations of simple probability, into various forms, often quite complex. However, any good method has a common-sense basis which can be made clear without much mathematical discussion. The term statistics has at least two meanings, which has sometimes led to harmful misunderstandings. It is often applied to mere tabulation, addition, averaging, etc. Statistical analysis of the type needed to interpret experimental results is far different from this clerical labor. I like Miss Gertrude Cox's term "experimental-statistics" and the former Civil Service classification "analytical statistics" which clarify the situation a little. "Biometrics" is a related term, applied to use of mathematical methods in biology, whether statistics or pure mathematics.

We should keep planning of tests to the fore in our study, since it is in planning that statistical methods find their fullest application. Past results give some basis for plans, and planned experiments are not only easier to analyze, but they get more out of the work and expense involved. In studying some old-time experiments in entomology, one is impressed by the "lost motion" and confusion as to results caused by want of good planning and analysis. In some cases intuitive conclusions were correct, in others different workers got conflicting results and arguments followed. Less con-

fusion and more progress would probably have resulted if good statistical practices had been available and had been applied.

The application of the methods we are discussing to biological experiments will usually point out conclusions that can be seen after analysis to be the common-sense ones. Progress is thus assisted, but is steady, slow, and unexciting. Occasionally statistical methods will make possible rapid progress in a single stroke, but this is not common.

An example of this sudden rapid progress was published some years ago (Holloway and Young, Jour. Econ. Entomol. 36(3): 453–457, 1943). Early workers in a certain area had jumped to a conclusion that an orchard scale insect of the region was checked by a fungus growing on the scales. The fact that scales increased following fungicidal sprays strengthened the belief. Restrictions on spraying and practices of broadcasting the fungus followed. Other facts led to suspicion that the rough spray residue, rather than the fungicide, favored the scales. A single well-planned and replicated test included:

1. A fungicidal, residue-producing spray;
2. A residue spray without fungicide;
3. A soluble fungicide which did not produce residue;
4. No spray at all.

Results showed conclusively that the residue, not the fungicide, produced scale buildup. Scale population was not affected much by the presence or absence of the fungus, which apparently grew mostly on dead or aged scales. Thus expensive practices of long standing were shown to be of little use.

CHAPTER 1

Survey of Problems and Methods

1. Nature and range of problems

In considering applications of statistical methods to entomological problems involving quantitative data, we must ask to what extent general methods will apply to our special problems. The answer is that they will apply well and be helpful in many places, if we take the nature of our data into account.

The problems of economic entomology are largely those of insect populations. Two phases of the problems recur constantly; density of insect population per unit, and proportion of the population affected by some condition. These are problems of *enumeration* of indivisible units. Problems of *measurements* are also encountered; weather records, spray deposits, crop yields, size of individuals, time required for a given process. The enumerations, in small samples, do not behave statistically exactly like measurements. Classic statistical methods are based on behavior of "continuous" variables such as measurements. A little modification in procedure is sometimes needed with enumeration statistics. Where numbers are large or counts are quite similar, standard methods are fairly accurate. Where high and low counts are mixed or numbers are small, it sometimes gives better results to *transform* the counts before analysis to some such function as the logarithm or square root.

It should always be remembered that the material we study is only a sample, and we study it to find what we can about the whole *population* of interest. Effect of some practice on corn borer population in field plots gives an idea of its effect on corn borers of the entire region. Successive samples will not agree exactly. Their variation is used to give a key to how near they are to the true population mean.

Statistical methods are finding considerable application in cage or laboratory experiments, field population studies, field plot experiments on insect control or on populations, and various other experiments such as those with livestock insects and insect traps. Considerations of sampling and experimental design applying to such studies will be shown in later chapters.

Cage and laboratory tests are usually on percentage of insects killed (an enumeration), but are sometimes on insect reactions, physiology, etc.

The writer (Wadley 1943a)* has set out some considerations in such tests. Favorable cage results are commonly tested on a broader scale in the field. Replication should be provided in laboratory tests, and if possible a full set of tests should be run on each day.

Field plot studies are common, especially on insect control. They often involve cooperation with agronomists, pathologists, or others, and there are several measures of results. Effect of practices on insect population is of major interest. Evaluating results by crop yield alone runs the risk of missing important points, though yields are needed for study. Especially superficial is the effort to state benefits of treatment in *dollars;* such estimates are of little value to future work, and partake more of extension than research. Correlation of insect density with crop loss is a subject needing more attention than it has received. Insect populations on plots are usually estimated by sampling.

Among methods of data collection, direct counts of insects in their habitat are foremost. Insect traps have been used as a sort of unsteady index of ups and downs of insect populations, as have sweep-net studies. Such indices often given results useful in immediate problems, but relations to true population density should be established if trap or net studies are to have the greatest usefulness in knowledge of populations. Great caution should govern utilization of such records, as they vary both from true population changes and from variable insect behavior. Morris (1960) has made valuable contributions in this field. Efforts have been made to reduce insect population to a low point by trapping. Scott and Milam (1943) record one, and later studies have been made. The fluid character of an insect population mobile enough for trapping makes such control very difficult, though it is a very plausible method. Fleming et al (1940) note experiments comparing traps. The "capture-recapture" method used by students of vertebrates has been applied in entomology in the case of tsetse fly populations. Estimates are made by releasing marked individuals, trapping, and estimating totals by the proportion of marked flies in the catch (Jackson 1940 and earlier papers). Variance of such estimates is high. Extent of injury to crops or animals is a measure definitely related to insect populations.

Experiments with insects attacking domestic animals have received some attention in studies of insecticides and repellents. Fryer et al (1943) have reviewed some methods and described experiments. Extensive series of mosquito repellent studies have been described by Travis et al (1949) and noted by the writer (Wadley 1946). Plant resistance to insects (Painter 1951) and infectivity of insect carriers of plant and animal disease are experimental subjects of importance. Peto (1953) gives some mathe-

* Refers to Literature cited at the end of the chapter.

matical methods for the latter case. Work of recent years in eradication of the screwworm is still being studied for broader applications.

A large variety of important work is carried on to estimate insect populations over wide areas by sampling. Such studies are usually termed surveys. Figures are used to compare with other times or other areas, to forecast need for control, or to check effectiveness of control work. In some cases, the question is merely whether or not a species is present at all; more frequently, the density of population is to be estimated. Both objectives are useful, but should not be confused (Wadley 1945a). If objectives can be defined, an intensity of sampling sufficient to detect any important population can be set up, and work can thus be limited. Surveys are not as adequate a basis for decision as experimental studies in improvement of control methods.

In some other types of problems, only a beginning has been made in use of biometrics, although pure mathematics has been applied in some cases. These problems include those in old-line life history studies, and ecological studies such as population growth curves, temperature-development studies, host-parasite relations, life table studies, and bioclimatics.

Studies of life history and habits were characteristic of the older work in entomology; numerous individuals were reared under rather restricted conditions, and time required for different stages was recorded. We now recognize the limitations of such studies, but they are still needed in preliminary stages of certain problems. Records are adaptable to biometric study; analysis of past studies would probably show that often more were reared than was necessary.

The subject matter developed under the heading of animal ecology (Chapman 1931, Allee 1949, Andrewartha and Birch 1954) is of broader application and great importance in attack on many entomological problems. Both physical factors such as temperature and biotic factors, such as association, competition, and population growth, are included. Volterra (see appendix to Chapman 1931), has given a theoretical mathematical treatment of population growth. DeBach and Smith (1941) developed a study of oscillation in host-parasite relations, with a useful review of literature. Waters and Henson (1959) and Morris (1960) are among those contributing to recent advances. Many other articles have appeared. Cook (1929) discussed bioclimatic problems. He stated the principle that the study of regional abundance and climate should go side by side with that of relation of time changes in abundance and weather. He distinguished carefully between economic and "taxonomic" distribution. Bioclimatic studies are hindered by lack of accurate population records, but in some problems precise sampling methods are making records available.

In all such studies, biometric methods are needed to measure and account for variation, and to test the reliability of estimates evaluations. The writer, with associates, has made some such studies (Wadley 1936, Wadley and Wolfenbarger 1944).

Applications of biometrics to insect physiology and taxonomy are so far very restricted, although pure mathematics is much used in physiology. Insect physiology and behavior are very important in many problems. Wigglesworth (1961) and Roeder (1953) have given us two general texts with many references. Most work in physiology has so far been on a rather narrow basis, emphasizing careful studies under closely controlled conditions. Exposure to more varied conditions is usually necessary before application of findings can be made. There is a temptation to try to explain biological happenings solely from chemical and physical considerations, which may prove a snare. As with ecological studies, application of biometrics is needed.

Insect behavior is a neglected field of considerable importance, allied to physiology. The extreme variability of insect reactions and the anthropocentric viewpoint of some older studies have been hindrances. Insects are not capable of as complex behavior as are vertebrates. Insects vary greatly with conditions and species. Olfactory responses seem more variable than optical or tactile reactions. Biometric principles will help in plans and interpretation of tests, but experimental difficulties must first be attacked. Wellington (1957) has discussed some of these problems.

Taxonomic problems differ from some others. In setting up groups as distinct species, characters are sought which will distinguish them clearly without overlapping other species. With similar groups, sampling should be widespread, among various lots of each group. If a character is qualitative, and determinable by inspection without measurement, and it occurs without fail in 100 specimens examined, we can be fairly sure it does not fail in over 5 percent of the whole population. (This is based on the "Poisson" series to be mentioned later.) If a quantitative character or measurement is used, 20 or 30 well-distributed individuals will give fairly good estimates of the mean, and standard deviation (discussed later). In the whole population, practically all individual measurements will fall within 3 standard deviations of the mean.

Where two or more characters are measured, the best combination can be found by an application of multiple regression methods to secure the so-called "discriminant function" (Fisher 1946 and later editions). An example will be presented later.

In taxonomy in general, we wish to be able to separate all individuals of one group from those of another, without overlapping; while in many biometric studies it is sufficient merely to show that group measure means

are different. Simple applications of biometric methods in taxonomy are scattered through the literature. A few examples are in articles by Mickel (1928), Emerson (1935), Forbes (1953), Buchanan (1947). The last-named article does not discuss statistical methods, but they were used in defining the ideas developed. At present new methods involving cytology, biochemistry, and other techniques are being used in taxonomy, and revolutionary changes are developing. Articles by White (1957) and Micks et al (1966) will show this trend.

2. Methods of computation. Distributions

Methods of computation are to be presented for very simple problems illustrating the principal techniques of analysis. It is hoped that such illustrations will lead to wider use of the many excellent textbooks now available, some of which are cited at the end of this chapter. The examples will seem very elementary to those already analyzing their own data, and still more so to statisticians. They may be helpful to those with no experience in analysis, and to those with mathematical experience who do not know the simplest and most efficient statistical methods.

The idea of *distribution* of a *variable* quantity is basic in statistical analysis. In its simplest form it is only a grouping to show the order present in apparently jumbled observations. For example, 53 individuals of an insect species were reared at a certain temperature. The days required for development are listed as 10, 10, 10, 11, 12, 12, 12, 9, 10, etc. When arranged in a frequency table we have:

Days	No. of insects
9	1
10	10
11	19
12	21
13	1
14	1

Note that records are in whole days, since examinations were made only once a day. This "discontinuity" does not prevent the calculation of useful measures, especially where numbers are ample.

From such a table may be calculated measures of magnitude (the arithmetic mean is usually the best), and measures of spread such as the range and the standard deviation. Such a distribution may readily be expressed graphically. If we have enough information, we can describe the distribution mathematically and fit a curve to it. This has been done with several well-known distributions. A graph (fig. 1) will further illustrate the idea.

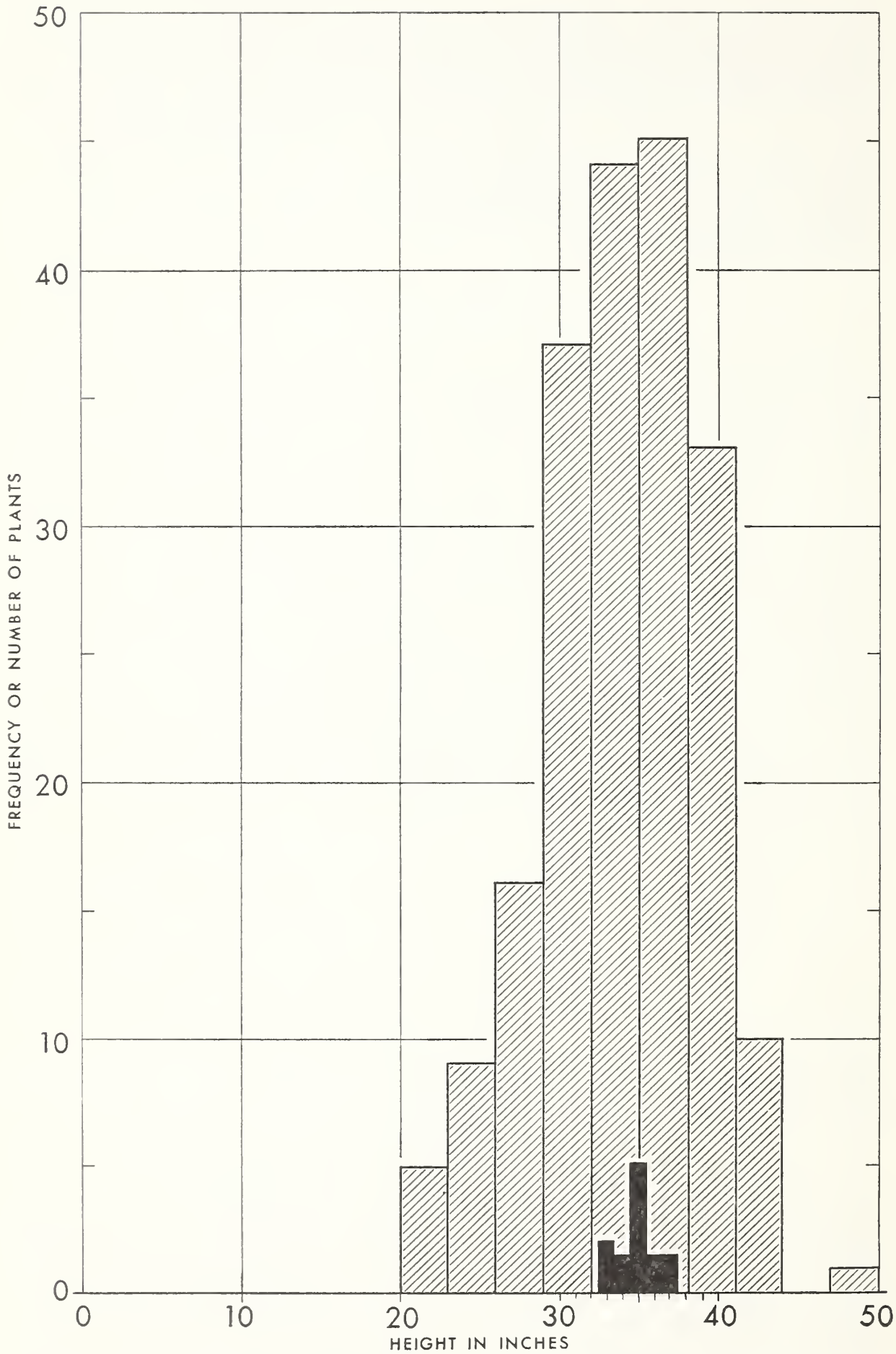

FIGURE 1. Histogram of heights of 200 corn plants. Ten means of 20 plants each are indicated in heavy marking.

Our statistical tests in most cases rest on knowing something about the possible distribution of results. For example, with a difference of two means (say *mean yields* of two crop varieties) we can calculate how much apparent difference might show up, just from chance variation, if there were no real difference. If the actual difference found exceeds this possible difference, we say the difference mentioned is real or "significant." To estimate the possible difference mentioned, we need some idea of the variation, gained by having several repetitions of the experiment, and of the underlying mathematical distribution of differences in repeated tests of the same thing.

A well-known distribution long studied by mathematicians is the "normal" distribution. It is symmetrical and is determined mathematically by the mean and standard deviation. It is fully described in texts to be cited; tables have been made of areas under the curve corresponding to lengths measured along the base. These tables can be used in tests. Some individual values (especially measurements) tend to fall in this distribution when successive determinations are made; the "days required" above would probably do so. *Means* of a series of values, from repeated random samples, are almost sure to be normal in distribution; as are differences of two means. Many standard statistical tests are based on the normal distribution.

A "statistic" in the truest sense is an estimate from a sample of one of the basic constants of a population, such as its mean or standard deviation. When the sample is small, such estimates may not be very good. A modification of the normal distribution, called the "t" distribution, is used for testing in such cases. It approaches the normal when numbers are large. Practically all our tests involving the standard error will be made in the t distribution. The tests presuppose that individuals were randomly sampled from the population (see Chapter 2), and the individuals were randomly distributed in space or time.

The *binomial* distribution is applicable to numbers or proportions of individuals having or lacking a given characteristic. It can be defined by expansion of a binomial, $(p + q)^n$, where p is the proportion of successes, q is $(1 - p)$, and n is the number of cases. A simple example is given by the distribution of heads and tails in shaking up 3 coins a number of times. Here p is $\frac{1}{2}$ or 0.5, q is 0.5 and n is 3. The binomial is $(0.5 + 0.5)^3$. Expanding, we have $p^3 + 3p^2q + 3q^2p + q^3$, which with p and q each 0.5 becomes $0.125 + 0.375 + 0.375 + 0.125$. Accordingly, the expected results in a number of shakes is: 3 heads $\frac{1}{8}$ of the time, 2 heads and a tail $\frac{3}{8}$ of the time, a head and 2 tails $\frac{3}{8}$ of the shakes; 3 tails in $\frac{1}{8}$ of the trials. The distribution is obviously not continuous or uninterrupted like the normal, but has values at only a few points. With large numbers it becomes like the normal, but with small numbers and p or q near 1.0, it is

quite lopsided or "skew." It is the expected distribution for successive counts of percentage mortality and similar statistics, and is hence important to biologists.

Another "discontinuous" distribution, related to the binomial with p small, is the Poisson series of numbers per unit. The Poisson occurs when an organism has an equal chance of being in any of the units; and units have an equal chance of receiving an organism, unchanged by the presence or absence of organisms already in the unit. If m is the mean number per unit, and X is a series of small whole numbers (0, 1, 2, etc.), the Poisson distribution of expected proportion of zeros, ones, etc., is given by $e^{-m}m^x/X!$, which is equal to $m^x/(e^m)(X!)$. Here e is a mathematician's constant (about 2.72) and $X!$ is x factorial. Fitting will be shown in a later part of this work. The Poisson is the expected distribution for numbers of organisms per unit, such as insects per square foot or per plant. It is found, however, only where the organisms are *random* in distribution. In most natural populations a more complex modification such as the "negative binomial" will be found to fit better than the Poisson. Examples of this will also be shown later. Waters and Henson (1959) and the writer (Wadley 1950) have discussed applications of some of these distributions.

3. Methods of computation. Simple variation

The Fisher-Snedecor notation will be used in showing some examples; it is simpler than others, and especially easy for typing.

First will be shown a certain measurement on a series of individuals of two invertebrate species.

Sp. A: 7.6, 7.7, 7.8, 8.4, 9.3, 7.6, 7.0, 8.7, 7.0, 7.6
Sp. B: 9.4, 10.5, 10.0, 10.2, 10.1, 9.4, 9.6, 11.2

Taking first Species A, we denote the original measurement by X, the mean by $\bar{x}$. To get the variance and its square root, the standard deviation s, we need the *sum* of *squares* of *deviations* from the mean (table 1). Each deviation is $X - \bar{x}$; or conveniently, simply small x. The number of cases is denoted by n; here, $n = 10$. $S(\ )$ indicates summation.

We find $\bar{x} = S(X)/n = 78.7/10 = 7.87$. The calculation of sum of squares of deviations follows as shown. Then the standard deviation is calculated as $\sqrt{S(X - x)^2/(n - 1)}$, which is $\sqrt{4.7810/9} = \sqrt{0.5312} =$ about 0.73.

The sum of squares of deviations can also be calculated as $S(X^2) - [S(X)]^2/n$, which is more convenient with a machine. The term $[S(X)]^2/n$ is sometimes called the "correction factor." Here this calculation is $624.15 - (78.7)^2/10 = 624.15 - 619.37 = 4.78$, using only two decimal places.

TABLE 1. Calculation of sum of squares of deviations

X	$X - \bar{x}$	$(X - \bar{x})^2$	X^2
7.6	−0.27	0.0729	57.76
7.7	−0.17	0.0289	59.29
7.8	−0.07	0.0049	60.84
8.4	+0.53	0.2809	70.56
9.3	+1.43	2.0449	86.49
7.6	−0.27	0.0729	57.76
7.0	−0.87	0.7569	49.00
8.7	+0.83	0.6889	75.69
7.0	−0.87	0.7569	49.00
7.6	−0.27	0.0729	57.76
78.7	0.00	4.7810	624.15

TABLE 2. Sum of squares calculated using frequency

X	X^2	f(frequency)	fX	fX^2
9	81	1	9	81
10	100	10	100	1000
11	121	19	209	2299
12	144	21	252	3024
13	169	1	13	169
14	196	1	14	196
		53	597	6769

The sum of squares of deviations is very important in more advanced statistical work. Where several items have the same value, this fact may be utilized in calculation. The 53 lengths of time required for development, previously cited, may be treated as in table 2.

$$S(x^2) = S(fX^2) - [S(fX)]^2/S(f)$$
$$= 6769 - [(597)^2/53] = 44.3$$

The *variance* (V), the square of the standard deviation, is also much used in more advanced work. In table 1 it is 0.5312, calculated on the way to the standard deviation. The latter statistic is a steppingstone in the present problem of comparing means of Species A and B, but has some importance of its own. Where individual items follow a "normal" distribution, about two-thirds of the items are within one standard deviation of the mean (as, 7.87 ± 0.73); about 95 % within 2s; practically all within 3s. Individual variation is often important, as in taxonomic problems (Chapter 4, example 32).

If numerous samples of 10 were taken and a series of means were secured, they would show more limited variation than individual values. While individual items may not follow the normal distribution, means are almost sure to do so. The standard deviation of means, $s_{\bar{x}}$, can be estimated as $s/\sqrt{n}$. In table 1, this is $0.73/\sqrt{10}$ or 0.23. Thus estimated, $s_{\bar{x}}$ is usually called the standard error. Following up the characteristics of the normal distribution, when cases used to estimate s are numerous, two-thirds of the means of samples of 10 will be within one $s_{\bar{x}}$ of the true mean; 95 % will be within $2s_{\bar{x}}$. With our limited numbers ($n - 1 = 9$) the multiplier from the "t" distribution is 2.26 instead of 2; 2.26×0.23 gives 0.52.

This means that we are fairly sure that our mean, 7.87, is within 0.52 of the true mean; thus, that the true mean is between 7.35 and 8.39. There are the so-called fiducial or confidence limits. The calculation may use variances: $V\bar{x} = V/n$, $s_{\bar{x}} = \sqrt{V\bar{x}}$.

The comparison of the two series is a further development. The 8 individuals of Species B are used in calculations similar to those above, with results as follows: $S(X) = 80.4$; $\bar{x} = 10.05$; $S(X^2) = 810.62$; $S(x^2) = 810.62 - (80.4)^2/8 = 810.62 - 808.02 = 2.60$. Variance and standard error are quickly calculated. The difference of 2 independent means tends to be normally distributed with standard error (s_d) estimated as $\sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2}$. In the complete formula there is an allowance for correlation not important here. Textbooks cited later will show this. In problems such as this the best way of combining the variance is to calculate an average variance ("pooled") by adding the sums of squares of deviations and the quantities ($n - 1$). The latter quantities are known as the number of "degrees of freedom," a term better defined in textbooks, and to be referred to later in more detail.

The combination is carried out as in table 3. The variance of the difference is estimated as $V(1/n1 + 1/n2)$, or $0.4612\ (1/10 + 1/8) = 0.4612 \times 0.225 = 0.1038$. The standard error of the difference s_d is $\sqrt{0.1038}$ = about 0.32. The mean difference is $10.05 - 7.87$, or 2.18. The difference is about 7 times its standard error. The total of degrees of freedom, which fixes the reliability of the test, is 16 (table 3). Consulting the "t" table

TABLE 3. Pooling variance

Species	Sx^2	Degrees of freedom	Variance
A	4.78	9	—
B	2.60	7	—
	7.38	16	7.38/16 = 0.4612

TABLE 4. Yield of two varieties in 5 years

Variety	Year				
	1st	2nd	3rd	4th	5th
A	40	47	34	18	27
B	31	46	31	21	16
A–B	+9	+1	+3	−3	+11

at the end of this Chapter (table 9) we find the "5% point" for t with 16 degrees of freedom is 2.12; the 1% point, 2.92. For 16 d.f. (degrees of freedom) with no real difference, d/s_d will reach 2.12 only one time in 20 (5% point), and 2.92 only 1 time in 100. In our case $d/s_d = 2.18/0.32$, or 6.81. Hence we decide that the mean difference 2.18 could hardly belong to a population of differences with mean 0 and s_d 0.32. It must be real. If these samples of 10 and 8 are representative of their populations, the two species differ in this measurement. We would decide the difference to be real if computed $t(d/s_d)$ reached 2.12 or higher.

Note that this does not mean that odds are more than 19 to 1 in favor of the difference being real. Rather, the odds are less than 1 in 20 favoring getting such a large sample difference where there is no real difference.

In some cases items in two series are appropriately "paired" and a simpler procedure will work. A classic example is the yield of two varieties of wheat, grown side by side under the same conditions during 5 years (table 4). Pairing will eliminate the large differences between years, and will make s_d much smaller. Here the differences can be treated as a simple series: $\bar{x}$, $S(x^2)$, V, $V\bar{x}$, and $s_{\bar{x}}$ can be calculated in the "A-B" row. The "t" computed is $\bar{x}/s_{\bar{x}}$ and it may be compared with the tabular t for 4 d.f., to see if $\bar{x}$ is significantly different from zero.

$$S(X) = 9 + 1 + 3 - 3 + 11 = 21; \bar{x} = 21/5 = 4.2$$

$$S(X^2) = 81 + 1 + 9 + 9 + 121 = 221; S(x^2) = 221 - [(21)^2/5] = 132.8$$

$$V = 132.8/4 = 33.2; V\bar{x} = 33.2/5 = 6.64; s_{\bar{x}} = \sqrt{6.64} = \text{about } 2.6$$

Then t computed = 4.2/2.6 or about 1.6. With only 4 d.f., the 5% point is about 2.8. We conclude that we are not sure the varieties really differ, because of the large variation and small numbers. A longer series might or might not reveal a difference; on the other hand, the result we see might easily come about with no real difference.

Observe that if the procedure of table 3 were used, the error estimate would be higher because of the large variation between years. Subtraction of a correlation allowance (previously mentioned) would reconcile the

difference. Note also that pairing gave a possibility of better analysis, and that it would not have been possible with data of table 1. In that case the items were *unordered* without any appropriate order in their groups. Whether or not items are ordered is a very important factor in analysis. To introduce pairing or cross-classification where it does not belong may give weird results; to neglect it where it is appropriate sacrifices accuracy in analysis.

4. Methods of computation. Compound variation and analysis of variance

Variation is seldom as simple as in the examples above. We generally have more than one source of variation, and usually have several classes to compare instead of two. Statisticians have gradually perfected a technique which can be used if desired for simple problems and can be extended to more complex cases. The history of the struggle with the problem of compound variation and of the development of solutions, must be passed over here. The most convenient short-cut techniques will be illustrated with the data already used, and an indication given of possibility of extension.

For the measurement on Species A and B already discussed, the total sum of squares of deviations is first computed by throwing the two series together. The computation is: $[(7.6)^2 + (7.7)^2 + \cdots + (11.2)^2] - (78.7 + 80.4)^2/18$, or $1434.77 - 1406.27 = 28.50$. That part of this sum caused by between-species difference is estimated as $(78.7)^2/10 + (80.4)^2/8 - 1406.27$, or $619.37 + 808.02 - 1406.27$, which is 21.12. The sum of squares between items within classes could be computed in each class (as has in fact already been done; see table 3). However, since the sums of squares are additive, the within-class sum can be conveniently secured by subtraction. Thus we have $28.50 - 21.12$ which gives 7.38 (as in table 3). The analysis may be summarized (table 5).

The appropriate test here is the "F" test or variance-ratio test. Computed F is 21.12/0.46 or about 46.0. The distribution of F (where no real difference exists) is a different distribution for each combination of degrees of freedom. An abridged table (table 10) is given at the end of this chapter. Consulting table 10, we find, for 1 d.f. for greater mean square, 16 d.f. for

TABLE 5. Summary of analysis of variance

Source of variation	Degrees of freedom	Sum of squares	Mean square of variance
Total	17	28.50	—
Between species	1	21.12	21.12
Within species	16	7.38	0.46

lesser, that the 5% point is only about 4.5. If there is no real difference, F will go as high as 4.5, only 1 time in 20. Our value, 46.0, is far beyond this; we conclude as before that the difference is real.

From the calculations above, we can see that this procedure could be carried out just as handily with 3, 5, or 10 species, as with 2. The degrees of freedom between classes would be 2 for 3 groups, 4 for 5, etc. Those within classes could be summed for the various groups as is done above. The within-class or within-species mean square can be used to calculate a standard error of the difference between any two group means. This calculation for the problem above has already been done following table 3. Suppose there were 3 groups, each of 10 individuals; with means 10.05, 7.87, and 8.10, and variance within groups of 0.46. The standard error of any one of the 3 differences is estimated as $\sqrt{0.46(\frac{1}{10} + \frac{1}{10})}$ or $\sqrt{0.092}$ or about 0.30. This would show that the second and third mean did not differ from each other significantly, but that both differed from the first mean. The "F" test would merely show that a real difference between groups existed somewhere. The use of the standard error and "t" test as above, as a supplementary procedure, helps to pick out the cause of significance. In such a case the degrees of freedom within classes would number 27, for "t" or "F" tests.

Several assumptions underlie such analysis. It is apparent that we suppose that class means may differ, but internal class variances are similar and may be put together. To violate this assumption too far may lead to contradictory results, although the procedure is rather flexible. The similarity of variances is rather more important in the calculation and use of a standard error than in the "F" test. The items are supposed to be somewhere near normal in distribution within each class, and our enumeration statistics sometimes give trouble in this respect. In some cases, a function of the count numbers can be used in analysis ("transformation") which gets away from the trouble. These cases are discussed in another section. The need of transformation is really not pressing except in extreme cases; percentage counts near 0 or 100%, and highly variable population counts such as insect trap records, are examples. Additivity and independence of variances, as well as independence of means, is important.

Another caution must be observed in making "t" tests from the analysis of variance. Some differences may pass the 5% point merely because of the number of possible differences. With 10 means, there are 45 differences, and a couple of the highest will be expected to pass the 5% point even if no real difference exists. Hence, "t" tests should not be carried out unless "F" is significant. Special procedures for "t" tests in analysis of variance, allowing for numbers of classes, have been worked out. Snedecor (1956, Sec. 10.6) records one (see Chapter 4, example 33).

TABLE 6. Summary, analysis of variance of yields

Source of variation	Degrees of freedom	Sum of squares	Mean square
Total	9	1081	
Years	4	970	242.5
Varieties	1	44	44.0
Interaction: year × variety	4	67	16.8

A more complex and highly important case is that in which data are cross-classified. The data used in table 4 give a simple example. The sum of all 10 yields is 311; the 2 variety sums are 166, 145; the 5 yearly sums are 71, 93, 65, 39, 43. The total sum of squares of deviations is: $(40)^2 + (47)^2 + \cdots + (16)^2 - (311)^2/10$, which is 10,753 − 9672 = 1081 (table 6). The sum of squares between varieties is $[(166)^2 + (145)^2]/5 - (311)^2/10$, or 44. The sum of squares between years is $[(71)^2 + \cdots + (43)^2]/2 - 9672$, or 970. The remainder sum, 1081 − 44 − 970 = 67, the part associated with differential effect or "interaction" of varieties and years.

In this case, the interaction variance functions as error variance, and F for varieties is calculated as 44.0/16.8 = 2.62, not significant with 1 and 4 d.f. (see table 10). This confirms the "t" test of table 4, but the method of table 6 can easily be extended to 3 or more varieties.

As before, a standard error of differences can be calculated from the error variance, as $\sqrt{16.8\ (\frac{1}{5} + \frac{1}{5})} = \sqrt{6.72}$ = about 2.6, just as with table 4. Due to the manner of calculation, the "Y × V" variance is just half the within-class variance of table 4, and divisors equalize the variances.

In carrying out an analysis of variance, careful attention must be paid to whether or not there is "ordering" within classes, making cross-classification possible. In unordered classifications, the division of sum of squares is merely into "between" and "within." In ordered groups, two sources of variation and their interaction can be separated. Much more complex classifications than those shown can be analyzed. Where they are cross-classified 3 or more ways, special tables of subtotals can be arranged, and sums of squares can be separated as already shown. In some complex analyses, some classifications are ordered and some unordered (Cassil, Wadley, and Dean 1943). The textbooks cited, and examples to be shown in a later chapter, will indicate methods for more complex cases.

Another important question in analysis of variance is, what is the error variance expressing random variation? In complex analyses, we sometimes have two or more variances furnishing error for different questions (as in "split-plot" experiments). A variance from experimental treatment may be subdivided; we may have variety, fertilizer, and variety × fertilizer. These "treatment" interactions are computed just as is the interaction

in table 6, but do not function as error. The interaction of table 6 is an interaction between *controlled* (variety) variation, and *uncontrolled* (year) variation, and thus has an element of random error. The interaction of variety and fertilizer does not.

As indicated after table 3, analyses will show if there are differences in the figures available; *conclusions* must rest on what the figures represent. To draw conclusions about the *populations*, we must know the limited figures we have to be good *samples* of these populations.

5. Methods of computation. Correlation or association of variables

Often we have two varying factors occurring in each individual case, in such a way that their relation can be studied. An intuitive method of studying such a relation is that of *grouping*. For example, taking the July rainfall and the average corn yield in a given locality over a series of years, we can take all years with 3.1 to 4.0 inches of July rain, take the average yield for these years, and compare it with the average yield for years with 4.1 to 5.0 inches, 2.1 to 3.0 inches, etc. A second intuitive method is plotting both variables against time or order of occurrence, to see if ups and downs coincide.

A more fruitful approach is a graphic one such as portrayed in figure 2. The trend of association shows up in a way leading easily to further study. It is a short step to represent the trend by a line or curve, which may be concisely described by an equation. *Regression* means the change in one variable associated with the change in another, while *correlation* in the narrow sense is a measure of the closeness of the relation.

In figure 2A, population density per unit area of an insect pest is plotted on one axis, percentage of parasitism by its insect enemies on the other. The figures are shown in the first two columns of table 7. When the axes are marked off in convenient intervals, we take the first case (density 24 %, parasitism 24 %) and find place 24 units to the right of the zero point, 24 units above. At this location, we place a dot. The next dot will be 21 units over, 20 up. When all are placed, the dots give a picture of the relation.

It is a handy convention to place the variable we regard as causal ("independent") on the horizontal axis, the one showing effect ("dependent") on the vertical axis, and to call them X and Y respectively. Sometimes, of course we find variables that do not qualify as dependent, but are interchangeable, such as length and fineness in cotton fiber. Even here in the mathematical solution it is often necessary to treat one variable as dependent. With interchangeable variables we can have two regressions, Y on X and X on Y.

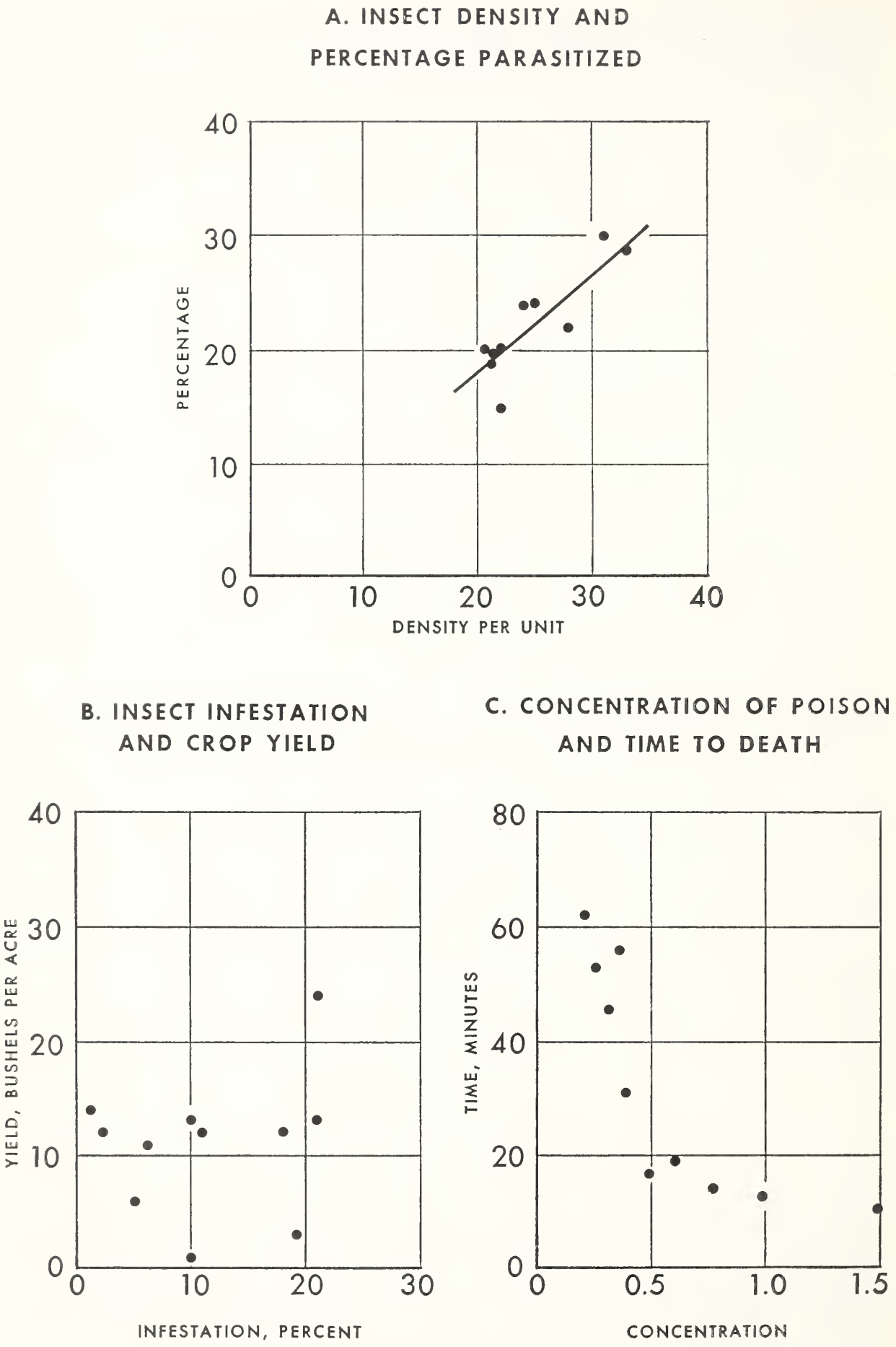

FIGURE 2. Preliminary plotting in correlation studies

TABLE 7. Insect density and parasitism

Location	Density (% of infestation) (X)	% parasitism (Y)	X^2	XY	Y^2
1	24	24	576	576	576
2	21	20	441	420	400
3	31	30	961	930	900
4	22	20	484	440	400
5	33	29	1089	957	841
6	25	24	625	600	576
7	22	15	484	330	225
8	21	20	441	420	400
9	21	19	441	399	361
10	28	22	784	616	484
Total	248	223	6326	5688	5163

In figure 2A the trend is plainly upward and about straight. In figure 2B data that have no trend are shown; in 2C a trend is evident, but it is curved and downward (negative). Preliminary plotting is the first important step in a two-variable regression problem. It will show us if there is a relation; if it is close or loose, positive or negative, straight or curved. It quickly shows the general relation and makes more exact study easier to plan.

In a number of problems a straight line or linear relation is adapted, or is close enough to the fact to be useful. Lines or curves may be fitted by eye or in several mathematical ways. The preferred line or curve is one which will reduce to the lowest possible point the *sum of squares of deviations* of actual cases from the line. This is the "least square" principle. We may try it with data of table 7. Here we need the sums of squares of deviations of both X and Y from their means, calculated as in previous examples. We also need the sums of the products, $S[(X - \bar{x})\ (Y - \bar{y})]$, or $S(xy)$, calculated with correction factor, $[S(X)]\ [S(Y)]/n$. The calculations are illustrated in table 7 and below.

Any straight line can be drawn if we have (1) its slope, (2) its position. The slope is given by the "regression coefficient" (usually called b), which is the average change in Y for one unit change in X. The position is determined by the fact that the least square line will pass through the means, or the $\bar{x}$, $\bar{y}$ point. The least square value for b is $S(xy)/S(x^2)$. When all values have equal weight, as is true in simpler cases, the regression equation is $Y - \bar{y} = b(X - \bar{x})$, but it can be put into the convenient form $Y = a + bX$, defining a as $\bar{y} - b\bar{x}$.

$$S(x^2) = 6326 - (248)^2/10 = 6326 - 6150 = 176$$

$$S(y^2) = 5163 - (223)^2/10 = 5163 - 4973 = 190$$

$$S(xy) = 5688 - (248 \times 223)/10 = 5688 - 5530 = +158$$

Note that the sums of squares of deviations cannot ever be negative, but the sum of products of deviations could be negative; hence the latter is shown with sign.

Then $b = S(xy)/S(x^2) = +158/176 = +0.898$; $Y = a + bX$; $a = \bar{y} - b\bar{x} = 22.30 - (0.898 \times 24.8) = +0.03$ and $Y = +0.03 + 0.898X$.

Having the regression equation, expected values of Y corresponding to any of X in the range 21–33 can be calculated. (The calculation should not be carried far beyond the actual range under study.) For $X = 21$, the expected Y is $+0.03 + (0.898 \times 21)$ or 18.89; for $X = 30$, it is 26.97. Marking in these two values in figure 2A, the line can be drawn through them, among the dots for actual cases.

The deviations of any Y values from the calculated one, can readily be taken; for example, for $X = 21$, $Y -$ (Y expected) is $20 - 18.89$. However, the sum of squares of deviations from the line can be secured by a short-cut calculation, $S(y^2) - [S(xy)]^2/S(x^2)$. The first term is of course the total sum of squares. The second is the part accounted for by the relation. If all 10 expected values were calculated for Y and the sum of squares of deviations taken, it would be equal to $[S(xy)]^2/S(x^2)$. The difference in the terms is the sum of squares of deviations from the line. It is the same as would be calculated by squaring and summing the 10 deviations from the line. Thus the sum of squares is divided into two parts; one explained by the relation, the other is the divergence from the relation.

The sum of squares of deviations from the line is here calculated as $190 - (158)^2/176$ which is $190 - 142$ or 48. Allowing for regression reduces variation in Y from 190 to 48. A standard deviation can be calculated around the line instead of from the mean of Y. It is the so-called standard error of estimate, s_{yx}. The degrees of freedom are here $n - 2$, because both mean and regression coefficient enter into the estimation. The value of s_{yx} in this case is $\sqrt{48/8} = \sqrt{6} =$ about 2.45.

The correlation coefficient "r" is estimated as $S(xy)/\sqrt{S(x^2) \cdot S(y^2)}$, or it may be calculated from the relation: $r^2 =$ proportion of variation accounted for, or 142/190. By either equation r is about $+0.865$. A perfect correlation would be 1.0; no correlation at all, zero. If the relation is negative (downward slope); $S(xy)$, b, and r will come out with minus signs. Thus r measures closeness of the relation on a standard scale, from -1.0 to $+1.0$, while b measures the relation in terms of the dependent variable. The r^2 is sometimes called the *coefficient of determination.*

Correlation coefficients should not be calculated where X is controlled or put in artificially. An example might be several selected concentrations of a poison used to bring about mortality. Regression procedure, however, may be used in such cases.

The standard error of b is estimated as $s_{yx}/\sqrt{S(x^2)}$, here $2.45/\sqrt{176}$ or about 0.185. In comparing b with zero, "t" is 0.898/0.185 or nearly 5.

This is well past the 1% point for 8 degrees of freedom, so that regression is undoubtedly real.

Significance of r is harder to test, and some old formulas given are incorrect, but the textbooks cited should make the test clear. Snedecor has a table of significance for r's of different numbers of degrees of freedom. At any rate, if b is significant, r is sure to be. Using standard errors of b and $\bar{y}$, the whole regression may be evaluated as to error. This is also explained in textbooks cited.

Thus is explained procedure for the simplest case of regression. Often two or more independent variables may be involved. One might wish to investigate the influence of both July rainfall and July temperature on corn yield over a series of years. This situation is hard to represent graphically, but equations are not difficult if straight-line relations can be assumed. For two independent variables, X_1 and X_2, and dependent variable Y, the six sums and products of deviations are calculated. The equation is: $Y = a + b_1X_1 + b_2X_2$. The b's are secured from the equations:

$$S(x_1^2)b_1 + S(x_1x_2)b_2 = S(x_1y)$$

$$S(x_1x_2)b_1 + S(x_2^2b_2) = S(x_2y)$$

The numerical sums and products of deviations are written in, and solution gives the b's. More complex problems follow the same logic. This is multiple linear regression.

It is very frequent in biology to find curved relations. Curves may be fitted in numerous ways. Two ways of using the precise methods above occur. First, one or both variables may be *transformed* to some function which gives a linear relation. Often the use of logarithms will accomplish this. The straight line may be fitted as above to the transformed function, and calculated values if retransformed to original units will give the fitted curve. In figure 2C, a fairly straight line is secured if reciprocals of time to death are plotted against logs of concentration. If only the independent variable is changed, the estimate is a least square one. However, if the dependent variable is changed, the estimate of the transformed variable is least square but the estimate of the untransformed one is not exactly least square.

The second method of handling curves is to take some power or powers of X as additional independent variables, and use multiple regression as above. For example, X may be called "X_1" and X^2 treated as "X_2". This gives a least square estimate. The equation in such a case would be $Y = a + b_1X_1 + b_2X_2$, and it will yield a curve.

In some cases multiple and curvilinear regression techniques are combined.

6. Methods of computation. Chi-square

The statistic called chi-square (often written χ^2) is essentially a ratio of an actual variance to a theoretical expected variance. Its form is that of the actual sum of squares of deviations, divided by the expected variance. If there is no real difference between the two variances, the average value of the quotient is of course the number of degrees of freedom. If the actual variance is greater, the ratio is compared with the distribution table of the chi-square for the given number of degrees of freedom. If computed chi-square passes the 5 % point, it is judged to indicate a real difference.

The distribution is a special case of the "F" distribution (table 10). In the bottom line of the table, for n d.f. and ∞ (infinite) d.f., if the F is multiplied by n it gives the chi-square for n d.f.

Chi-square can thus be used where we have an idea of *expected* variance to compare. All its various uses are of this nature. In most of our common chi-square tests, the actual numbers of individuals are of great importance; mere percentages cannot be used. It is less widely useful than analysis of variance; the theoretical variance is a minimum one, and where numbers are large, the chi-square often shows small unimportant differences to be significant. The actual computed variance derived in analysis of variance is usually more realistic.

The binomial and Poisson series already referred to have expected variances; pq/n for the binomial, where p is the percentage or proportion and q is $100 - p$ or $1 - p$. The expected variance is simply the mean ($\bar{x}$) for the Poisson. If we have a series of percentage counts from a population, each based on the same number, their sum of squares of deviations may be divided by the theoretical variance of the binomial to secure chi-square. Significance will show that the counts vary too much for the binomial; they are not homogeneous, and other causes than randomness are causing difference. The same sort of test based on the Poisson, may be applied to population counts of units of the same size. Examples will be shown later.

The most common use is an approximate one, the test of frequency distributions. The expected variance of numbers in each class (Poisson) is measured approximately by the numbers themselves. If O is the observed (actual) number in a class, C the calculated (expected) number by some theory, the contribution to chi-square is $(O - C)^2/C$. If these ratios are summed for each of a number of classes, the sum is chi-square for that problem. (Separate estimates of chi-square with their d.f.'s may be added and tested; the additive property is of value.) The smaller classes are usually combined to provide at least 5 expected per class, though smaller numbers are sometimes used.

TABLE 8. Spray residue on cabbage, chi-square test of normality

Range, residue per head	Number observed (O)	Number calculated on normal theory (C)	$(O - C)^2/C$
0.000–0.149	8	8.4	0.02
0.150–0.199	10	9.7	0.01
0.200–0.249	14	11.8	0.41
0.250–0.299	7	9.7	0.75
0.300–0.449	9	8.4	0.04
			1.23

The number of degrees of freedom is the final number of classes, minus the number of ways in which expected and actual must agree.

As an easy example, 48 heads of cabbage were examined for spray residue, and a normal distribution fitted to the results (table 8). The classes were at intervals of 0.05 units, but it was necessary to lump some at the ends.

Chi-square is 1.23. There are 5 classes after lumping. Expected and actual must agree in mean, standard deviation, and total number. This leaves 2 d.f. The 5% point (table 10) is 5.98, using $2 \times F$ for 2 and ∞ d.f. Chi-square is nowhere near significance; as far as we can tell from the limited sample, the distribution accords with the normal.

In the enumeration statistics so common in entomology, we have an idea of theoretical minimum variance (Poisson), and chi-square tests may be made. Some evidence may be secured as to whether two or more groups are really different. If their actual variance exceeds the theoretical, they may show significance by chi-square even if not really different. But if they do not differ significantly by chi-square, they will not differ by any broader test. Hence a quick application of chi-square, may sometimes save later work. Snedecor (1956) gives a number of formulas and examples of various uses of chi-square. Dosage-mortality procedures use chi-square. The short-cut formulas sometimes disguise the basic comparisons of variance; the beginner should pattern his calculations carefully after the published examples.

7. Methods of computation. Covariance analysis

This subject cannot be more than mentioned here. It is discussed in detail in texts by Snedecor, Goulden, and Fisher (see references), and examples will be shown in a later section. The subdivision of sums of squares in analysis of variance has been shown. Sums of products of deviations can be subdivided in a way exactly parallel to that used for sums of squares.

In some experiments, some independent variable not included in the

experimental plan (such as moisture content or pH) may be measured in each unit. We may then secure two sums of squares and one sum of products of deviations of dependent and independent variables. These sums may be subdivided according to the outline of the experiment. Correlation and regression may be calculated for total, for between classes, within classes, and even more complex subdivisions. Differences between classes may be evaluated after allowing for known regression effect, and results may thus be increased in value.

Covariance analysis may be seen to combine methods of analysis of variance and regression. It sounds complex, but after once starting analysis with a good guide such as those mentioned, no great difficulty should be found.

This is true of all the techniques mentioned. A start at calculation, with material in which the student is interested, will often lead to surprising progress in understanding. The discussion is closed with a recommendation to the student to "break the ice" by following out calculations such as above and in chapter 4, then trying the methods on other and similar problems.

A paragraph may be added on "non-parametric" tests, such as comparisons of ranking or signs. These are treated by Steel and Torrie (1960; see references) and many other writers. Such tests are labor saving in analysis, but are often less efficient in use of data than those methods described above.

8. Special problems in biology. Transformations

It has already been noted that statistics of biological problems are largely enumerations, which do not fall in a normal distribution unless numbers are large. The expected distribution for percentage counts is the binomial, with variance pq/n; for population counts it is the Poisson with variance equal to the mean. The variance in such cases is correlated with the mean; contrary to the supposition in analysis of variance that intra-class variances are similar and may be pooled. In natural populations, distributions are more likely to resemble the "negative binomial," with somewhat higher variance, but still with correlation of mean and variance.

To meet this situation, and allow analysis of variance free from worry about correlation, several statisticians have worked out transformations. A transformation is a function of the original observation, which is used in analysis instead of the original number. Conclusions are drawn from analysis on the transformed scale, though means and confidence limits may be transformed back to original terms for summarization.

Bartlett (1937) showed that use of square roots of population counts

was better in theory than use of the counts themselves in analysis. He suggested use of $\sqrt{X + 0.5}$ for small numbers; this gives a definite value to a zero count. Bliss (1938) published the "angle" transformation for percentage counts; it is $\sin^{-1}\sqrt{p}$, and is tabled in Snedecor's text (1956). Cochran (1938) reviews these situations and points out the use of logarithms as a transformation for measurement data varying widely from class to class, also for very variable population counts. Williams (1937) used the log of $(X + 1)$ for highly variable insect trap data, with marked improvement.

These three are the principal transformations in use, but some modifications of them, or new transformations, have been proposed. Upholt (1942), Beall (1942), and Fryer et al (1943) expressed dissatisfaction with existing transformations; the two latter articles proposed new ones. A transformation may be worked out for a specific problem, using methods outlined by Beall (1942).

On the whole, transformations are not of much help unless the need for them is extreme. Milne (1943) showed with sheep-tick counts that transformed and untransformed figures gave similar results in analysis. Williams's highly variable trap counts are an example of a situation where a transformation is needed. Another example is an analysis of percentage mortalities, with some classes near 50 % and some near 100 %. In such extreme cases, transformation will give more trustworthy results, and will often show differences not to be detected by analysis without transformation. In more ordinary cases, analysis of enumerations with and without transformation will often give closely equivalent results. The writer (Wadley 1943b) has proposed some standards for decision as to when to transform percentage counts. Percentages for analysis without transformation should be based on adequate and similar numbers; should be between 10 and 90 %; and should, if possible, have over 20 of both individuals succeeding and failing.

9. Special problems in biology. Probit analysis and biological assay

A considerable literature has grown up around the dosage-mortality curve, especially since Bliss's articles (1935). Finney (1952) has summed up much of this. It concerns an application of regression, to the relation of percentage mortality in insects and other organisms, to increasing concentration of a toxicant. The curve is S shaped, with the lower bend very small ("asymmetric sigmoid"). If logs of concentration are used, the curve becomes symmetric; if percentages of mortality are converted to standard deviation values or "probits," something like a linear relation results. Bliss and others have developed intensive methods of fitting a

regression line to mathematically weighted values, estimating concentration needed for 50 % (or other mortality), and comparing poisons through these estimates. The logical basis for comparison is the ratio of concentrations (or difference of their logs) needed for a given effect. The percentage mortalities have expected variances, which give a basis for testing

TABLE 9. Distribution points of "t"

Degrees of freedom	"t" for probability level of		
	50%	5%	1%
2	—	4.30	9.92
3		3.18	5.84
4	0.74	2.78	4.60
6		2.45	3.71
8		2.31	3.36
10	0.70	2.23	3.17
12		2.18	3.06
15		2.13	2.95
20	0.69	2.09	2.85
40		2.02	2.70
100		1.98	2.63
∞	0.6745	1.96	2.58

This table is adapted by permission from *Statistical Methods*, fifth edition, by George W. Snedecor, copyright 1956 by the Iowa State University Press.

TABLE 10. Abridged table of F, 5%

Degrees of freedom for lesser mean square	Degrees of freedom for greater mean square										
	1	2	3	4	5	6	8	10	15	20	100
1	161	200	216	225	230	234	239	242	246	248	253
2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.39	19.42	19.44	19.49
3	10.13	9.55	9.28	9.12	9.01	8.94	8.84	8.78	8.70	8.66	8.56
4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.96	5.86	5.80	5.66
5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.74	4.62	4.56	4.40
6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.06	3.94	3.87	3.71
8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.34	3.22	3.15	2.98
10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.97	2.84	2.77	2.59
12	4.75	3.88	3.49	3.26	3.11	3.00	2.85	2.76	2.62	2.54	2.35
15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.55	2.41	2.33	2.12
20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.35	2.20	2.12	1.90
50	4.03	3.18	2.79	2.56	2.40	2.29	2.13	2.02	1.88	1.78	1.52
100	3.94	3.09	2.70	2.46	2.30	2.19	2.03	1.92	1.77	1.68	1.39
∞	3.84	2.99	2.60	2.37	2.21	2.09	1.94	1.83	1.66	1.57	1.24
χ^2	3.84	5.98	7.80	9.48	11.05	12.54	15.52	18.30	24.90	31.40	—

For only a few degrees of freedom, 1% F is several times as large as the 5% values; for 5 and 5, about double; for 15 and 15, it is about 50% greater.

This table is adapted by permission from *Statistical Methods*, fifth edition, by George W. Snedecor, copyright 1956 by the Iowa State University Press.

results by chi-square. Examples of calculation will be shown in a later chapter.

Joint effect of insecticides has been discussed by Bliss (1939), Finney (1952), and others. Where a combination of insecticides gives an effect greater than could be expected from addition of their effects alone, we have *synergism.* This is discussed by the authors mentioned; the writer (Wadley 1945) gives short-cut methods for estimating synergism.

When a measurement of reaction in each individual can be used, rather than the proportion killed, study follows more general methods. Concentration is often expressed as logarithms, giving a fairly linear relation. This *graded response* is often used with vertebrate subjects, but not often with insects. Comparison of reagents by their effect on organisms, whether of percentage affected or of graded response, is called *biological assay.*

Campbell and Sullivan (1938) and Steiner (1939) discuss some older insecticide methods of value. Stomach poisons pose especially hard problems in giving uniform doses.

10. References

Out of a host of good textbooks available, a few may be cited which have been useful to the writer. Those selected have clear and practical approaches not too hard for the working scientist.

Ezekiel, Mordecai, and Fox, Karl A.
1959. Methods of correlation and regression analysis; linear and curvilinear. 3rd ed., 548 pp. Wiley, New York.

Good in correlation studies. Shows technical derivations in appendix. The authors are interested in economics, but approach is broad enough to be helpful to biologists.

Finney, D. J.
1963. Probit analysis. Rev. ed., 318 pp. Cambridge U. P., New York.

An excellent compendium on this special subject, important in entomology.

Fisher, Ronald A.
1958. Statistical methods for research workers. 13th ed., 356 pp. Hafner, New York.

This work has been the foundation of great advances in the application of statistics in experimental work. All the many editions are good; the later ones have more material. The writer uses the 10th edition, 1946. Fisher's rather difficult style is easier after studying Snedecor and Goulden.

Goulden, Cyril H.
1952. Methods of statistical analysis. 2nd ed., 476 pp. Wiley, New York.

An excellent text full of experimental viewpoint, following Fisher's line of thought.

Hoel, Paul G.
1962. Introduction to mathematical statistics. 3rd ed., 253 pp. Wiley, New York.

Well described by the title.

Smith, C. A. B.
1966. Biomathematics. 4th ed. 2 vols. vol. 1 Algebra, geometry, calculus. 1965. vol. 2 Numerical methods, matrices, probability, statistics. In prep. Hafner, New York.

Helpful for a simple approach to pure mathematics.

Snedecor, George W.
1956. Statistical methods. 5th ed., 534 pp. Iowa State, Iowa.

This text has had a profound influence on American science in introducing Fisher's great advances in a moderately "painless" way. Earlier editions are somewhat simpler and easier than the fifth.

Steel, Robert G. D., and Torrie, James H.
1960. Principles and procedures of statistics. 481 pp. McGraw, New York.

A new text of considerable value giving sound principles, good examples, and recent advances in ideas.

Walker, Helen M.
1951. Mathematics essential for elementary statistics. Rev. ed., 248 pp. Holt, Rinehart & Winston, New York.

A self-teaching manual.

Yule, G. U., and Kendall, M. G.
1958. An introduction to the theory of statistics. 14th ed., 552 pp. Hafner, New York.

This text has a long history of usefulness, and in a way connects mathematical and practical statistics. Older editions had many references to original work; present editions have fewer references.

11. Literature cited

Allee, W. C.
1949. Principles of animal ecology. 837 pp. Saunders, Philadelphia.

Andrewartha, H. G., and Birch, L. C.
1954. Distribution and abundance of animals. 782 pp. Univ. of Chicago, Chicago.

Bartlett, M. S.
1937. The square root transformation in the analysis of variance. Suppl. Jour. Roy. Stat. Soc. 3: 68–78.

Beall, G.
1942. The transformation of data from entomological experiments so that the analysis of variance becomes applicable. Biometrika 32: 243–262.

Bliss, C. I.
1935. The calculation of the dosage-mortality curve. Ann. Appl. Biol. 22: 134–167.

———
1935. The comparison of dosage-mortality data. Ann. Appl. Biol. 22: 307–333.

———
1938. The transformation of percentages for use in the analysis of variance. Ohio Jour. Sci. 38: 9–12.

———
1939. The toxicity of poisons applied jointly. Ann. Appl. Biol. 26: 585–615.

Buchanan, L. L.
1947. A correction and two new races in Graphognathus(white-fringed beetles). Jour. Wash. Acad. Sci. 37(1): 19–22.
Campbell, F. L., and Sullivan, W. N.
1938. Testing fly sprays. Soap and San. Chem. 14(6): 119–125, 149.
Cassil, C. C., Wadley, F. M., and Dean, F. P.
1943. Sampling studies on orchard spray residues in the Pacific Northwest. Jour. Econ. Entomol. 36: 227–231.
Chapman, R. N.
1931. Animal ecology. 448 pp. McGraw-Hill, New York.
Cochran, W. G.
1938. Some difficulties in the statistical analysis of replicated experiments. Empire Jour. Exp. Agr. 6: 157–175.
Cook, W. C.
1929. A bioclimatic zonation for studying the economic distribution of injurious insects. Ecology 10: 282–293.
DeBach, P., and Smith, H. S.
1941. Are population oscillations inherent in the host-parasite relations? Ecology 22: 363–369.
Emerson, A. E.
1935. Termitophile distribution and quantitative characters as indicators of physiological speciation in British Guiana termites(Isoptera). Ann. Entomol. Soc. Amer. 28: 369–395.
Fleming, W. E., Burgess, E. D., and Maines, W. W.
1940. Relation of color to the attractiveness of Japanese beetle traps. Jour. Econ. Entomol. 33: 320–327.
Forbes, W. T. M.
1953. Note on multimodal curves. Ann. Entomol. Soc. Amer. 46: 221–224.
Fryer, F. C., Shaw, A. C., Atkeson, F. W., Smith, R. C., and Borgmann, A. R.
1943. Technique for conducting fly repellency tests on cattle. Jour. Econ. Entomol. 36: 33–44.
Jackson, C. H. N.
1940. The analysis of a tsetse fly population. Ann. Engen. 10: 332–369.
Mickel, C. E.
1928. Biological and taxonomic investigations on the Mutillid wasps. U. S. Nat. Mus. Bul. 143. 351 pp.
Micks, D. W., Rehmet, Anne, and Jennings, Judy.
1966. Biochemical differentiation of morphologically indistinguishable strains of *Aedes aegypti*(Diptera:Culicidae). Ann. Ent. Soc. Amer. 59: 239–241.
Milne, A.
1943. The comparison of sheep-tick populations (Ixodes ricinus). Ann. Appl. Biol. 30: 240–250.
Morris, R. F.
1960. Sampling insect populations. Ann. Rev. Entomol. 5: 243–264.
Painter, R. H.
1951. Insect resistance in crop plants. 520 pp. Macmillan, New York.
Peto, S.
1953. A dose-response equation for the invasion of micro-organisms. Biometrics 9: 320–335.

Roeder, K. D.
1953. Insect physiology. 1100 pp. Wiley, New York.
Scott, L. B., and Milam, J.
1943. Isoamyl salicylate as an attractant for horn worm moths. Jour. Econ. Entomol. 36: 712–715.
Steiner, L. F.
1939. The laboratory-field method for testing codling moth insecticides. U. S. Dept. Agr. E-488, mimeographed.
Travis, B. V., Morton, F. A., Jones, H. A., and Robinson, J. H.
1949. The more effective mosquito repellents tested at the Orlando, Florida laboratory, 1942–47. Jour. Econ. Entomol. 42: 686–694.
Upholt, W. M.
1942. The use of square root transformations and analysis of variance with contagious distributions. Jour. Econ. Entomol. 35: 536–543.
Wadley, F. M.
1936. Development-temperature correlation in the green bug, *Toxoptera graminum*. Jour. Ag. Res. 53: 259–266.
———
1943a. Statistical aspects of laboratory tests of insecticides. Amer. Assoc. Adv. Sci., Publ. 20., pp. 177–188.
———
1943b. Statistical treatment of percentage counts. Science, 98: 536–538.
———
1945. On the application of the Poisson series to some problems of enumerations. Jour. Amer. Stat. Assoc. 40: 85–92.
———
1946. Incomplete block designs applied to paired tests of mosquito repellents. Biometrics Bul. 2: 30–31.
———
1950. Notes on the form of distribution of insect and plant populations. Ann. Entomol. Soc. Amer. 43: 581–586.
——— and Wolfenbarger, D. O.
1944. Regression of insect density on distance from center of dispersion as shown by a study of the smaller European elm bark beetle. Jour. Agr. Res. 69: 299–308.
Waters, W. E., and Henson, W. R.
1959. Some sampling attributes of the negative binomial distribution, with special reference to forest insects. Forest Sci. 5: 397–412.
Wellington, W. G.
1957. Individual differences as a factor in population dynamics; the development of a problem. Canad. Jour. Zool. 35: 293–323.
White, M. J. D.
1957. Cytogenetics and systematic entomology. Ann. Rev. Entomol. 2: 71–90.
Wigglesworth, V. B.
1961. The principles of insect physiology. 5th ed., 546 pp. Methuen, London; Dutton, New York.
Williams, C. B.
1937. The use of logarithms in the interpretation of certain entomological problems. Ann. Appl. Biol. 24: 404–414.

CHAPTER 2

Sampling

A sample is defined as a portion of a population taken for study, in the hope that it will be representative enough to tell us what we need to know about the whole population.

Entomologists usually show great interest in sampling. The necessity of making decisions about uncounted millions of insects from observation of a few thousand or a few hundred makes entomologists conscious of the need for sound sampling. For example, the writer once reared 53 insects of a certain species at 17°C., to estimate time required for development. The 53 constituted a sample of the entire species, and the writer hoped to draw conclusions about the species from them.

The writer has occasionally heard the remark (thankfully, seldom from entomologists) "we are interested only in the actual material studied." This is a very illogical point of view, which can rob experimental science of its meaning. We are interested in the material studied, for what it can tell us of the whole population.

Often we study insect infestation in a plot or field by sampling, when the field itself is a sample of the larger population investigated. Thus we have samples within samples.

In some cases, as in quarantine problems, it is important only to see if the species is present at all or not. Then sampling is simply a matter of looking intensively in the likeliest places. In the more usual and more important cases, we are interested in the *density* of population and good sampling methods can improve results and reduce work. In entomology we practically never develop estimates of the *total* population of a field or region, such as are developed by Census people and some economists. We could develop such estimates, but in practice we always estimate densities or rates of infestation. These seem more sensible and helpful in our insect populations. The number of corn borers per 100 stalks is a good example. In general, dense populations vary more absolutely, but less proportionately, than do sparse ones.

1. Essentials of sampling

In sampling we wish to attain *representativeness* to get as accurate a picture as possible; to avoid *bias*, which is a tendency to err persistently in one direction; to consider *randomness*.

We judge representativeness partly by reproducibility; if repeated samples give closely similar results, or if *sampling error* is low, we believe our samples to be representative. This condition is best achieved by getting the sample from as many parts of the population as possible. For example, if fruit in an orchard is to be judged from a sample of 1,000 apples, it is better to take 50 apples from each of 20 well-distributed trees, than to take 250 apples from each of 4 trees. It might be still better to take 10 apples from each of 100 trees, but the time required for collection may be excessive. The limitation on spreading sampling is usually a practical one. Unskilled samplers sometimes think that a representative sample can be secured by "purposive" selection of units which they consider typical. Such sampling is unsafe and may lead to bias. The population to be sampled must be clearly defined.

Freedom from bias is a second and related principle. Objectiveness, or freedom from personal choice, is an important factor. A man sampling plants in a field, for example, to determine percentage infested or diseased, may find that his eye strays subconsciously to the sort of plants he is interested in. Thus his sample shows too high an infestation. If he notices this, he is apt to adopt some device for taking the choice out of his own hands. Years ago, some students of Hessian fly on wheat threw their trowels well out into a wheat field, then took a strip of drill row nearest the trowels for study. A series of plants would thus be taken in one place, and several places would be taken, with bias pretty well eliminated. Not all bias is personal. In a study of yield data in individual drill rows of wheat, it was once found that every eighth row was deficient in yield, probably because of a defect in the drill. A sampling scheme based on taking some of every eighth row would give badly biased results. Bias is serious, especially if unrecognized; a biased sample may be very reproducible and misleading.

Randomness may be defined, in a simple way, as giving every unit in the population an equal chance to appear in the sample. Representativeness and freedom from bias are needed for a good estimate of the population mean. Randomness is to insure a good estimate of sampling variation or error. Fisher (1960) makes clear that randomization is the basis of validity in the error estimate. Randomization is achieved by some system of drawing numbers or such devices (see Snedecor 1956 for tables of random numbers). As a simple example, suppose we are going to measure 20 corn stalks as a sample of height in a corn field. We draw a number to indicate which row to take, a second number to indicate the order of a plant in a row, and with these two numbers we have located a stalk at random. Repeating the process 20 times will make up the sample. In practice, sampling from field or nonfield material may not be subject to such elaborate procedure. We try to achieve a fairly wide distribution of units and to rule out personal choice,

and feel some confidence that choice is practically random. However, in cases where doubt is felt, a careful randomization is reassuring.

2. Interrelation of factors

A random sample is unlikely to be biased. Bias may be avoided, however, without randomness. Randomness and representativeness are somewhat in conflict. Often a more accurate estimate can come from a sample with more even distribution than a fully random sample. Randomness gives us a valid error estimate for the mean. The estimation of the mean itself is of course the primary object. If we desire merely to estimate conditions at a single time and place, we might even dispense with an error estimate altogether. But if comparisons are to be made with other times and places, as is usual, the error estimate is vital.

If the units of a sample are spaced absolutely evenly through the field, we have a *systematic* sample. The equivalent in laboratory material is to take every *n*th unit in regular order; when the starting point is once determined the sample is fixed. The systematic sample differs somewhat in its behavior from a random sample. Sometimes, in very uniform material, they behave alike. But in a field or population varying from part to part, the systematic sample will be more accurate than a random sample. On the other hand, if a variance and standard error of the mean is calculated in a systematic sample as if it were random, the standard error will be higher than it would with a random sample. Thus the systematic sample does not give a trustworthy error estimate; while the random sample yields a helpful error estimate. Efforts to calculate errors for systematic samples by special methods have not been very satisfactory. The even distribution gives an accurate estimate of the mean, and by including a maximum of variation, gives too high an error.

A useful and common practice is to compromise and *restrict* randomness, dividing the field into parts as uniform as possible within, and getting the major variation between parts. Then each part is sampled randomly. This preserves much of the accuracy of the systematic sample, and retains some randomness as a basis for error. This kind of restricted sampling is often called stratified random sampling. The variance *within* parts or strata is the basis for the standard error of the mean, which is usually lower than for a wholly random sample.

In speaking of measuring stalks in a field, the plan of measuring 20 plants each at a separate place was noted. In practice we would probably go to fewer spots in the field and take several stalks in each place, to save time. The variance between adjacent stalks, however, tends to be low and inadequate for an error estimate. The error should in such a case be based on spot

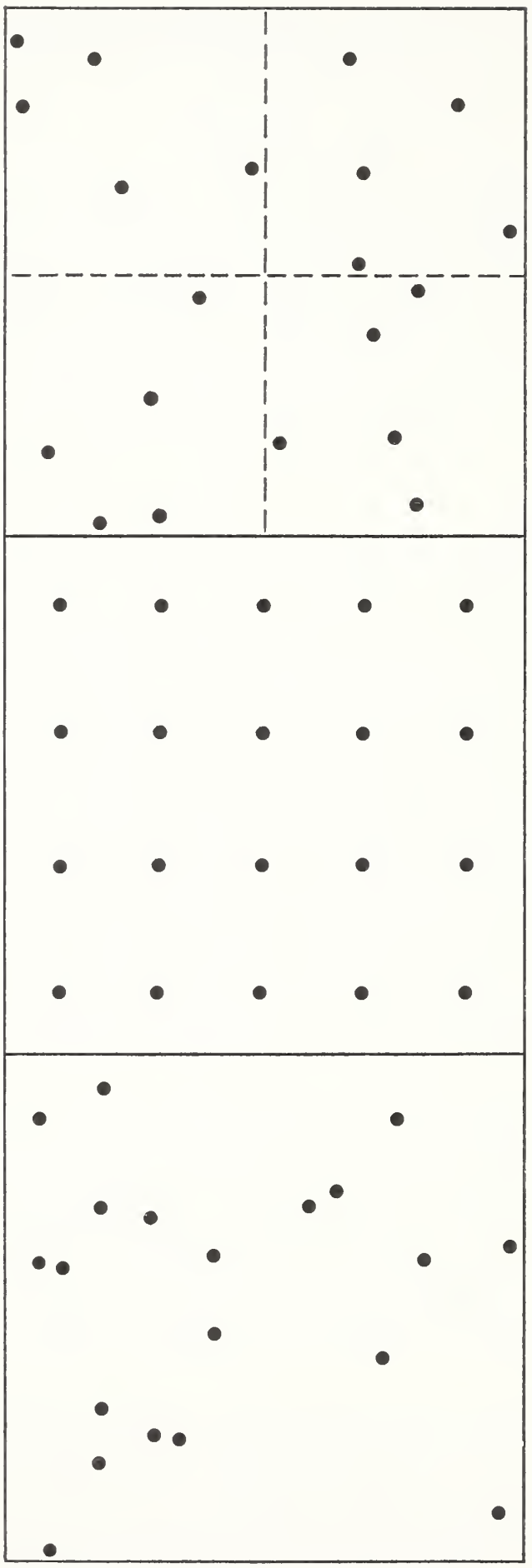

FIGURE 1. Three plans for sampling plants in a field

means; if stratified sampling is practiced, there should be several spots in a stratum. Even if 3 complete but separated rows in the field were taken, we would strictly speaking only have 3 units in the sample. Several sampling methods are diagrammed in figure 1.

3. Comparison of methods

A concrete example may be drawn from insect counts recorded by Fleming and Baker (1936); the writer had access to the detailed records. There were 2,500 individual square-foot counts of Japanese beetle larvae in a 50 × 50 foot area. Populations at that time were high and numbers rather uniform, so that it did not seem necessary to transform counts for analysis. Samples were taken by fully random, systematic, and stratified random plans, with 50 square-foot units per sample. This gave an approach to the actual field conditions under which insect populations must be estimated.

Five samples were taken by each method. The systematic samples were taken by dividing the area into 50 rectangles, each 5 × 10 units, and taking a unit at the same position in each rectangle. The starting point was changed for each complete sample. Purely random samples were taken by locating a row and column, randomly, to fix each unit. Stratified random samples were taken by dividing the area into 25 squares, each 10 × 10, and taking 2 units by random choice in each square.

The 5 purely random samples gave means and standard deviations as in table 1.

The standard error of the mean (averaged as squares) shows an average value of about 0.9. The standard deviation calculated empirically among the 5 means is a little over 1, with a fairly good agreement. This demonstrates roughly that a useful measure of reproducibility can be calculated from a single sample.

When the 5 systematic samples were used as if they were random, results were secured as in table 2.

The pooled standard deviation is 7.2, which will give an estimate of standard error of the mean of 50 a little over 1.0. Actually, however, the empirical standard deviation calculated among the 5 means is only about 0.5.

TABLE 1

	1	2	3	4	5
Mean	18.5	19.3	20.5	19.3	17.7
S.D.	6.3	5.9	7.6	5.5	6.4
$s_{\bar{x}}$	0.9	0.8	1.1	0.8	0.9

TABLE 2

	1	2	3	4	5
Mean	19.1	19.4	19.6	18.6	18.4
S.D.	6.8	7.3	7.1	7.5	7.0

This illustrates the tendency of standard deviations to run high, and means to be more reproducible, in systematic than in random samples. It shows that systematic samples are likely not to yield a trustworthy error estimate.

Each of the 5 stratified random samples was studied by analysis of variance. A typical analysis is shown in table 3.

The standard deviation of random sampling is estimated as $\sqrt{34.8}$ or 5.9; the standard error of a mean of 50 as $\sqrt{34.8/50}$ or about 0.8. For the 5 samples results are shown in table 4.

The pooled standard deviation is about 6, the estimate of $s_{\bar{x}}$ about 0.8, and the standard deviation of the 5 means computed about the true mean is about 0.7. This shows that the error estimate from any one sample is a useful measure.

The accuracy of the three methods is reflected in the standard deviation of sample means around the general mean; fully random 1.0, systematic 0.5, stratified random 0.7. Thus the stratified random method conserves accuracy and at the same time permits a useful error estimate. Accuracy may be improved by selecting the strata to be as uniform within (in respect to the measure studied) as possible, even though they are not of regular shape. In the example above regular size and shape of the strata were arbitrarily fixed.

TABLE 3

	Degrees of freedom	Sum of squares	Mean square
Between blocks	24	1946.5	81.1
Within blocks	25	869.5	34.8

TABLE 4

	1	2	3	4	5
Mean	19.2	20.3	19.4	18.5	18.9
S.D. (within stratum)	4.9	6.7	5.1	5.9	7.0
$s_{\bar{x}}$ estimated	0.7	0.9	0.7	0.8	1.0

4. Devices for improving sampling

The stratified random sampling plan is widely usable. When the number of units taken in each part is proportional to the size of the parts, the sample is self-weighting. Weighting will be discussed more fully later. A refined mathematical method of determining numbers for each subdivision is to make them jointly proportional to size and standard deviation where the latter is known; that is, proportional to the product of these quantities. Where all standard deviations are similar, the number of units taken in each part is proportional to the size of the part.

In much insect work such choice of size of subdivision sample may lead to oversampling of a part of the field that is large in size but small in importance, or to the reverse. Perhaps as good a method as any is to sample each subdivision as adequately as possible and combine the results if a general average is needed.

In one special situation the writer once suggested a sort of tandem sampling. With field populations of aphids, density varies enormously, from sparse populations hard to find to swarms literally coating plants. The suggestion was to make a preliminary view to find roughly the level of population, and later to use sampling methods for dense populations different from those for sparse populations. In this as in other situations, it is desirable to use standard methods that will mean the same in the hands of various samplers.

Subsampling or compound sampling is another device important and practical in many situations. The major sampling units are not completely studied, but data are determined by subsamples. A familiar illustration is that of estimating the wheat yield of an area by visiting a number of fields and estimating the yield of each one by a moderate-sized sample. We may regard experimental plots as units of a sample, and if the insect infestation of each plot is itself estimated by sampling, we have compound sampling. Analysis of variance applies conveniently to such cases, and by such analysis we can separate the effect of the major and minor orders of sampling on precision. The sampling variance of major units functions as error for questions based on these units; that of minor units within the large groups will be included in the major error.

In problems of broad scope compound sampling is more usual than simple sampling. In practical work it is not necessary to use randomness in locating minor units within major units, but it is necessary if the minor units are to be used in studying technique. Major units should have some element of randomness, as error estimates are based on them.

Another type of subsampling is the sampling in the laboratory of material gathered as a composite sample from the field. This practice is familiar to

chemists and other laboratory technicians. Henderson and McBurnie (1943) have described a method of this type of subsampling with mite populations on citrus leaves that reduced labor considerably. In the setting up of such a method, it must be shown that no bias is brought about. Bias may be avoided by using more laborious methods of known exactness as a standard of comparison. Where two or more subsamples are provided for each sample, subsampling variance may be determined.

Henderson and McBurnie also describe mechanical methods of mite collection. Such methods are frequently developed by workers. They are not directly statistical, but are an outgrowth of desire to get the most out of limited time and funds. Pielou (1957) has contributed something to theory in such problems. Sometimes insects can be weighed or measured if an efficient collection procedure is available, and if the relation of the quantity thus determined to actual numbers is established.

The last type of subsampling is a form of double sampling in which the characteristic of interest is hard to measure. We therefore estimate a related characteristic, easier to handle, on a good-size sample, and estimate the relation of the desired characteristic from a more limited sample, with a saving in labor. Double sampling takes various forms. In a study of European corn borer populations, the desired characteristic is borers per 100 plants, but counting is laborious, requiring careful dissection of stalks. Therefore, the percentage of stalks infected is easily estimated on a large sample, and a limited sample is dissected to determine the number of borers per infested stalk. The final figure is the product of these two. Cases might occur in which the final figure would be a quotient of two variables. In other cases regression of the first character on the second is estimated from a medium-sized sample, the second is then estimated from a large sample, and the final figure is computed from the estimated regression applied to the results of the large sample. This method could be applied when numbers of insects are estimated from measurement or weight.

Double sampling is useful where material has been placed in categories by rapid inspecting methods, as has sometimes been done with number of scale insects on citrus, or amount of damage by earworms to corn (Wadley 1949). If material in each category is sampled and the samples are used in actual counts, the means of these counts may be applied to the categories to form the estimates. Estimation can sometimes be considerably improved without a great increase in work. Determination of error in double sampling is complex. Where products or quotients are used, and are calculated separately in every replication or major subdivision, error may be simply calculated among the final figures.

An important development of recent years is *sequential* sampling. The sampling units are taken and examined, and the sample repeatedly evalu-

ated during the work until the information is sufficient to place the mean inside or outside of previously determined limits. Then work may stop. This will obviously often result in a saving in work. The method is adaptable to some surveys. Waters (1955) has had encouraging results with sequential sampling in forest insect surveys. His article describes procedure and gives mathematical methods and references.

In insect-population work some index of the population is often used rather than an actual count. Sometimes active or numerous insects are caught with a sweep net, instead of being counted on the plants. Trap catches or screen counts often serve as indices of abundance. Use of such methods assumes a correlation; the correlation must be established if they are to have any usefulness. Investigation sometimes shows that sweeping, for example, gives different results on windy and calm days, or that it gives an incorrect picture of sex ratio. These methods are often useful for immediate decisions, but correlation with exact populations must be established if they are to lead to real gains in the knowledge of insect populations.

One method of interpreting sampling results must be viewed with doubt. A sampler will sometimes array his data in classes by some qualitative criterion, assign rank numbers to the classes, and proceed to use these numbers as if they were measurements. If infestations are graded as 1, 2, and 3, for example, we have no assurance that 2 is twice as heavy as 1, or that 3 is as heavy as 1 and 2 put together; 2 may be five times as heavy as 1. Such analysis may result in loss of part of the value of an experiment or sampling study. Double sampling can be applied in this situation with profit.

5. Some special considerations

In sampling in entomology we are generally interested in density of population or in the proportion of the population affected by some characteristic. Both are determined by counting indivisible units, rather than by measuring. This means that sampling variance has a limiting value below which it cannot be expected to go. No amount of precision in procedure will make the variance lower than the minimum value; for percentage counts, the binomial variance pq/n; for population counts, the mean. The standard error of the mean may of course be reduced by taking larger samples. In low population densities the variance among units is usually close to the theoretical minimum, and in high densities it is greater in proportion to this theoretical value. In low populations sampling error is lower absolutely, and higher proportionally to the mean than in high populations. In percentage counts variance between successive counts is comparatively low near zero and 100 percent, and higher at intermediate values.

A special consideration is that of sampling from a limited population.

If the sample makes up a large part of the entire population, it approaches a census. If we measure every plant in a field, we know the average absolutely, without any sampling. If we measure 25 or 50 percent of the plants, the true standard error of the mean will be lower than the classic formula indicates. We can of course estimate the standard deviation accurately from such a large sample. If n is the number of units in the sample and N the number in the whole field of inquiry $s_{\bar{x}} = (s/\sqrt{n})\ \sqrt{1 - n/N}$. Using variances as more convenient, we may write $V_{\bar{x}} = (V/n)\ [1 - (n/N)]$. If n is small in proportion to N, this is the ordinary formula, since $1 - (n/N)$ is practically 1. Unless the sample is more than 10 percent of the whole, the adjustment is unimportant. It is of slight importance to entomologists, since our samples are usually small in proportion to our field of inquiry. One entomologist was sampling bark on a large tree for insect infestation and the units were extremely variable. He calculated the standard deviation from a number of units, and attempted to estimate how many units would be required for a desired low standard error, using the equation $s_{\bar{x}} = s/\sqrt{n}$. The answer was absurd, as it indicated that more units must be taken than existed on the tree. The equation $V_{\bar{x}} = (V/n)\ [1 - (n/N)]$ gave a reasonable answer. The idea of taking a certain percentage of the population in a sample, often seen in economics, does not apply well in entomology. The size of the sample, rather than its percentage, gives precision.

6. Weighting

Weighting in sampling results has caused considerable confusion. The basic principles are as follows: (a) If several parallel samples from the same material are to be combined, weighting by number of units in the sample is appropriate. (b) If the samples represent different parts of a field of inquiry, the best estimate of the average is obtained by weighting by the sizes of these parts. If the parts are equal or nearly so, no weighting is needed. It is assumed that each part is sampled fairly adequately. The mathematical principle of weighting by reciprocal of variance is involved; it has proved of value in some situations such as probit analysis, but need not be developed further here.

As an example of weighting, suppose that three samples, such as those discussed in the section on restricted randomness, are taken from the field, each representing all parts of the field. If one is of 50 units and the other are two of 100 each, they should be weighted accordingly. This can be accomplished by adding the totals and dividing by 250 for the mean per unit. If we wish to work with the means per unit, already calculated from the three samples, we can multiply the small sample mean by 50, each of

the other two means by 100, add the products, and divide by 250. If we keep the same proportions, we can simplify the multipliers to 5, 10, and 10 and divide by 25, or to 0.2, 0.4, and 0.4 without any division of the sum.

Suppose, however, that the field is of 40 acres, 16 of one soil type and 24 of another. In combining the two samples we give the mean of one part a weight of 16/40, the other a weight of 24/40. The result is our best estimate of the average condition in the entire field.

It is obvious that in the latter case we may be combining things not very similar, and that a more critical procedure would be to state the averages separately. However, we are constantly being called upon for statements of averages, such as the average crop yield for a State, or the average infestation of some insect for a county. Obviously, we will not always know the proper weights, and so must use approximate estimates or assume equality.

Snedecor (1956) discusses sampling and standard errors of weighted averages in more detail than can be done here. Where we have several parallel samples from the same material, we are really combining several samples into a single larger sample. The variance and standard error may be calculated as if it were one large sample. Variances, if already calculated, may be pooled by adding sums of squares of deviations and degrees of freedom, and dividing.

When several samples from different parts of the material are combined, Snedecor gives the formula for variance of a weighted average as

$$V_{\bar{x}w} = S[Vw^2/K]/[S(w)]^2$$

where V is the variance among individual units in each class, K is the class number, and w is the weight to be used in each. The principle is that, when variances from unlike classes are combined, they should be weighted by the squares of the class weights instead of being pooled as are those from like classes.

In one case of insect-population sampling, four environments with equal weight had 6 units each, and a fifth had 70 units, but was to be weighted by only 3.5 because of its small area, while the other environments had weights of 6 each. To obtain a weighted average, the statistics are as in table 5.

The weighted mean is $[(6.0 \times 8.3) + (6.0 \times 10.7) + \cdots + (3.5 \times 13.5)]/27.5$, or 8.1. The variance of this mean is the sum of the column Vw^2/K divided by the squared sum of the w's.

$$234/(27.5)^2 = 0.31$$

Extracting the square root, we obtain $s_{\bar{x}}$ as 0.56 (rounding to 0.6). The

Table 5

Environment	Mean	Number of units (K)	Weight (w)	Variance of units (V)	Vw^2/K	
A	8.3	6	6.0	13	$(36 \times 13)/6$	$= 78$
B	10.7	6	6.0	7	$(36 \times 7)/6$	$= 42$
C	7.0	6	6.0	11	$(36 \times 11)/6$	$= 66$
D	3.3	6	6.0	7	$(36 \times 7)/6$	$= 42$
E	13.5	70	3.5	35	$(12.25 \times 35)/70$	$= 6$
Sum	—	—	27.5	—		234

mean then is 8.1 ± 0.6. The large variation between environments does not enter the standard error here.

7. Planning and interpreting sampling

In planning a sampling study the first thing to consider is the objective. The information sought should be clearly defined. If we desire merely to record the presence or absence of an insect species, elaborate sampling suited to estimating population density will not be needed. We need only look carefully in likely places. This situation partakes more of quarantine than of research philosophy. If we desire to estimate density, however, looking in the likeliest places is almost sure to give too high an estimate. When sampling for density we must inspect both lightly and heavily infested places.

Next we must consider the methods to be used. We should keep in mind the factors of representativeness, freedom from bias, and randomness, with their functions. Efficiently planned sampling will give better figures for the same amount of work and expense, or equally good figures with less work, than poorly planned sampling. We should utilize all available previous information. Our object in quantitative sampling is, first, to estimate the average numbers, and second, to obtain an idea of the variability. If we have some preliminary idea of variability, we can estimate the amount of sampling needed for an estimate of given accuracy and precision. This accuracy can be measured as the standard error of the mean. In the equation $s_{\bar{x}} = s/\sqrt{n}$ we can supply a preliminary estimate of s, an acceptable figure for $s_{\bar{x}}$, and solve for the n. The differences measurable or likely to be missed by the sampling can also be defined. If no preliminary estimate of s is available, it is often wise to carry on some exploratory work to obtain one. In such work we will be sampling for the standard deviation or variance rather than the mean. With insect-population counts we may always have in mind the minimum theoretical standard deviation.

It is often possible to modify the plan of work midway in investigations,

if study of early results suggest methods of gaining efficiency. The precision (measured as $s_{\bar{x}}$) of determination of the mean is governed only by the size of the sample (n) and the variability (s). The percentage of the entire population in the sample has no great influence. Taking a fixed percentage is not a sound statistical procedure; a 5-percent sample is a better sample in a large population than in a small one.

Whether such devices as stratified sampling, compound sampling, or double sampling will be helpful depends on the nature of the problem. A knowledge of the material to be sampled will aid in efficient stratification. Arbitrary subdivisions can be made if there is no such knowledge, but more efficient work is usually possible if the subdivisions can be made along lines of known variation.

We may have fields within a district as our principal sample units, and small areas within fields as minor or subsample units. The variation of fields within a district is more important than that of units within fields. The degree to which each source of variation contributes to the error of the final results can be evaluated by use of analysis of variance.

A good example is the preharvest estimation of wheat yield in a county, by using 20 fields as units in a sample of the area, and well-distributed but small subsamples in each field. A small sample will give us nearly as good an idea of the yield in a field as a large one. Differences between fields will usually be larger than between units within fields. If we take a very large subsample, or even a complete harvest, of a few fields, we know the situation in those fields very well, but we do not know the county average well, because fields vary. If we take limited subsamples in each of a large number of fields, we get a better estimate of the county average for the same work. Grasshopper egg sampling (Davis and Wadley 1949) is a parallel case.

In such a setup the standard error of the county mean will be estimated by calculating the standard deviation between field means and dividing it by the square root of the number of fields taken in the county. This standard error will include the large field-to-field variation, and will also have a smaller component caused by sampling variation within fields. If sampling variation within fields is absent (if a complete harvest of each was taken), the standard error will be somewhat smaller. By the use of analysis of variance we can estimate the error due to within-field sampling, if within-field units as well as fields are taken randomly. The units can be stated in any convenient form, as yield per subsample unit or per acre, in pounds or in bushels. Suppose we have 20 fields and 5 units per field, with results as follows: Between fields—Degrees of freedom 19, Mean Square 89; within fields—d.f. 80, M.S. 29.

From this summary, using the mean square within fields as B, the mean

square between fields as $kA + B$, where k is the number per field (5), we can calculate A, the variance between fields over and above that within fields, on a unit basis. This is estimated as $(89 - 29)/5$, or 12. Variance of the mean for any combination of n fields and k units per field will be estimated as $A/n + B/nk$. In this case it will be $12/20 + 29/100$, or 0.89, and the standard error will be $\sqrt{0.89}$, or about 0.94. If we have 50 fields with 2 units per field, the expected variance of the county mean will be $12/50 + 29/100$, or 0.53, and the standard error about 0.73. For 10 fields and 10 units per field the standard error would be 1.22.

In this manner we can estimate the effect of changes in sampling plan. To spread out sampling will always give a gain if the mean square between fields exceeds significantly that within fields, and if A has a real existence, which is usually true.

The analysis shown can be adapted to the study of small adjacent areas within one field, and thus to comparison of a few large units with a larger number of smaller units. In such a comparison we think of the large units as made up of adjacent smaller units, and make our analysis within and between larger units. If the smaller units completely occupy the larger unit, they are essentially random; the random choice of the larger unit makes them so. By this method 4 or 5 spots in a field, with 25 units per spot, were found to give as precise results as 2 larger spots with 100 units per spot. In orchard sampling for results of spraying, 8 plots of 1 tree each gave as good results as 4 plots of 3 trees each, under conditions of the orchards used.

Labor and other costs must enter into sampling plans. Very fine subdivision of sampling will often greatly increase the labor of covering the ground. Examination of additional units in a spot may add little expense, and it will increase precision somewhat, even though not so much as studying more spots. We must figure, not the lowest standard error possible, but one that will be acceptably low and within our limits of work and expense. Often a compromise can be made and a good plan worked out that will hold down cost and provide for enough exactness. In an elaborate sampling investigation, however, it may pay to subject costs as well as variances to a more exact study. If we have the variance of large sample units and of subsample units, the cost of each type of unit, and the total allowable cost, we can solve for the best combination of n and k in the following equations:

$$V_{\bar{x}} = A/n + B/nk$$

$$T \text{ (total cost)} = nCD + nkC$$

Here C represents the direct cost of each subunit, CD the overhead cost of each major unit above subunit costs; C, CD, A, B, and T are fixed; and we

solve by calculus for n and k giving the lowest value for $V_{\bar{x}}$. Tippett (1940) treats this problem in his chapter on experiments. In the solution

$$k = \sqrt{\frac{BCD}{AC}} \text{ or } \sqrt{\frac{BD}{A}} \text{ and } n = T/(CD + kC).$$

For a fixed $V_{\bar{x}}$ and lowest total cost, k is the same, and n is estimated as $(kA + B)/kV_{\bar{x}}$ (Davis and Wadley 1949).

8. Literature cited and other references

Some of the books listed below contain good discussions of sampling principles. Snedecor's text contains much discussion of sampling, with the last chapter devoted to sampling principles. Cochran (1953) gives some recent advances, and there are other good recent texts. Much recent work is quite complex, and is largely concerned with questionnaire methods. Morris (1960) has published a valuable discussion of insect sampling with many references.

Literature Cited

Cochran, W. G.
1953. Sampling techniques. Pp. 330. Wiley, New York.
Davis, E. G., and Wadley, F. M.
1949. Grasshopper population in the northern Great Plains and its relation to egg survey methods. U. S. Dept. Agr. Circ. 816, 16 pp.
Fisher, R. A.
1960. The design of experiments. 7th ed., 248 pp. Oliver and Boyd, London.
Fleming, W. A., and Baker, F. E.
1936. A method for estimating populations of larvae of the Japanese beetle in the field. Jour. Agr. Res. 53: 319–331.
Henderson, C. F., and McBurnie, H. V.
1943. Sampling technique for determining populations of citrus red mite and its predators. U. S. Dept. Agr. Circ. 671, 11 pp.
Morris, R. E.
1960. Sampling insect populations. Ann. Rev. Entom. 5: 243–264.
Pielou, E. C.
1957. The effect of quadrat size on the estimation of the narameters of Neyman's and Thomas's distributions. Jour. Ecol. 45: 31–47.
Tippett, L. H. C.
1940. The methods of statistics. 3rd ed., 278 pp. London.
Snedecor, G. W.
1956. Statistical methods. 5th ed., 534 pp. Iowa State University Press, Ames.
Wadley, F. M.
1949. An application of double sampling in evaluating insect infestations. Jour. Econ. Entom. 42: 396.
Waters, W. E.
1955. Sequential sampling in insect surveys. Forest Sci. 1: 68–79.

Other References on Sampling, General and Entomological

Beall, G.
1939. Methods of estimating the population of insects in a field. Biometrika 30: 422–439.

Bliss, C. J.
1941. Statistical problems in estimating populations of Japanese beetle larvae. Jour. Econ. Entom. 34: 221–232.

Cassil, C. C., Wadley, F. M., and Dean, F. P.
1943. Sampling studies on orchard spray residues in the Pacific Northwest. Jour. Econ. Entom. 36: 227–231.

Cochran, W. G.
1939. The use of analysis of variance in enumeration by sampling. Jour. Amer. Stat. Assoc. 34: 492–510.

Henderson, C. F.
1960. A sampling technique for estimating populations of small arthropods in soil and vegetation. Jour. Econ. Entom. 53: 115–121.

Hendricks, W.
1942. The theory of sampling. U. S. Dept. Agr. and North Carolina State College. Pp. 123.

Jones, E. W.
1937. Practical field methods of sampling soil for wireworms. Jour. Agr. Res. 54: 123–134.

King, A. J., McCarthy, D. E., and McPeek, M:
1942. An objective method of sampling wheat fields to estimate production and quality of wheat. U. S. Dept. Agr. Tech. Bul. 814, 87 pp.

Ladell, W. R. S.
1938. Field experiments in the control of wireworms. Ann. Appl. Biol. 25: 341–389.

Larrimer, W. H., and Cartwright, W. B.
1926. Determination of the percentage of infestation by the Hessian fly, *Phytophaga destructor* Say. Jour. Agr. Res. 32: 1049–1051.

Meyers, M. T., and Patch, L. H.
1937. A statistical study of sampling in field surveys of the fall population of the European corn borer. Jour. Agr. Res. 55: 849–872.

CHAPTER 3

Experimental Design

1. Philosophy of experimentation

An experiment may be defined as a trial involving comparisons or testing theories. The term also carries with it the idea that conditions are arranged to some extent; that at least part of the extraneous variation is controlled, so that the things to be compared are different, and other conditions are the same.

An actual example may be drawn upon to illustrate the difference between experimentation and ordinary sampling. Suppose a small area of a forest to be treated by a certain method, and a second part left untreated. Insect populations on the two areas are to be compared by sampling. Now the samples are perfectly valid to compare two areas as such, or to compare the same area at two different times. The sampling may succeed in proving the two areas are different, but cannot furnish evidence that the treatment caused the difference. The experimenter will be unable to say whether the difference is caused by treatment, by characteristics of the two areas, or by a combination of both. To secure real evidence on the effect of treatment, *several* separate or independent areas are necessary.

The philosophy of experimentation needs more attention than it receives in many cases. Scientific proof, the place of fact and inference, of induction and deduction, will repay study. Some experimenters have an intuitive desire to explain away the variation in each individual case, and refine their procedure and results to an undue degree. This approach if carried too far may lead to results which other experimenters fail to duplicate, or to results close to previous expectation. Procedure should of course be precise, but after it is once launched results should be accepted by the experimenter. Hard-won data may be discarded if breakdown in experimental conditions is shown, but should not be discarded for statistical reasons alone, or because they point to unexpected conclusions. We have seen much more progress in experimental work; trying for undue refinement has been replaced by hopeful comparisons, which are made to see what differences appear in spite of variation. We must accept the fact that a certain amount of variation is normal in biology, and accept and utilize this variation to give an estimate of experimental error.

Thinking of experimental problems will lead to formation of hypotheses

or questions for test. The experiment should be set up to test these questions as positively as possible. The experimenter should make up his mind in advance what decision will result from each possible outcome of the experiment. After the data are in, he should stick to his decision. A doubtful decision may call for more trials, but failure to accept an unexpected result is undesirable. Such failures have held back progress at times.

Needs in experimentation are threefold. First, objectives should be clarified as in the foregoing discussion. Second, the experiment should be *valid*, or capable of giving real evidence on the question tested. The forest test mentioned above was lacking in validity. Third, the experiment should be as *efficient* as possible, so as to give maximum results for time and expense. If careful planning gives a desired result with ¾ of the work of a less well planned test, a nice gain has been made. Most printed discussions of experimental design deal largely with efficiency, taking for granted clarity of objectives and validity. However, we are not past the need for considering all three. The experimenter should endeavor to do *reproducible* work; the methods and criteria should be described so that they can be carried out by others.

2. Some principles of experimental design

We may state a useful classification of experiments as preliminary, critical, and extensive. Preliminary tests are limited in extent, and have the function of exploring methods or showing the behavior of unfamiliar material. Critical experiments are carefully planned, and aim to lead to final decisions if possible. Extensive experiments are meant to reduce proved principles to practice, and they may approach the commercial in scale. Biometric methods are helpful with all the classes, but are most essential with critical experiments.

Two things should be remembered in studying design methods. Experimental design has developed along with analysis of variance, which is the solution for compound variation, and the design used specifies the form of analysis of results. Design has largely grown up in field plot studies, with ever-present field variation, and has been adapted later to nonfield experiments.

Any self-contained critical experiment must provide for two things; a fair comparison of things or "treatments" tested, and an estimate of *error* of comparisons. This error can be defined as the variation which would be shown if there were no real difference between things tested. We can estimate it in a true sense, only by *replication*, or repetition of tests. In the hypothetical forest test mentioned, it is not provided for, but it would have been available if several areas had been treated. *Randomization* of

experimental treatments among experimental units is necessary to insure that the error estimate will be a good and valid one (Fisher 1956). The uncontrolled variation furnishes the error estimate. Sometimes it is a direct estimate; if treatments A and B are tested, the variation among the trials of A, and that among the trials of B, will give the error. Sometimes the data are cross-classified according to a controlled and an uncontrolled source of variation; then the differential effect or interaction furnishes the error estimate. By a controlled source we mean something reproducible such as a variety or a concentration.

Several principles besides replication and randomization are important in development of experimental designs. The first is that of *parallel comparisons*, comparing things side by side under as similar conditions as possible. Comparison of the yields of two varieties of wheat is much more precise, if they are compared side by side in the same years and the same fields, than if variety A is tested in several places, variety B in several other places. Comparison of several insecticides is much better if they are tested each time on the same days, in the same laboratory, and on samples from the same stock of insects. In planning the experiment, making these parallel comparisons corresponds to restrictions on randomization in sampling. Arrangement of such comparisons is the principal source of gain in efficiency.

Where several things are to be tested, the question will often arise as to how many can be put in one experiment. For example, if one or more insecticides are to be tested in field plots, several strengths of application, and several schedules of repeated application, may be used. A little study will show that fewer total experimental units or plots will be needed, if dosage is tested in one experiment, schedule in another. However, this will give no information about differential effect of dosage and schedule. An experiment having all combinations of all factors studied is called a "factorial" experiment. It will often be best in the long run, especially if differential effects are important. The differential effects or interactions between *controlled* factors are subdivisions of treatment effect. They are not like the interactions mentioned above which are used as error, although they are estimated arithmetically in the same way.

When two or more factors are mixed up in planning so that their effects are confused, they are said to be "confounded." In the forest test mentioned, the effect of treatment was confounded with that of area. Unintentional confounding is wasteful, and defeats the aim of experiments, but careful and discriminating confounding has some uses in advanced experimental design.

Plot size is a practical question of great importance in field experiments. The plots should be large enough to allow for expected border effect, and

to have an area in the center which will really reflect treatment conditions, and will supply enough material for study. Increasing size of plots tends to reduce error, but not as much as increasing their number. Small and numerous plots are thus more efficient than larger and fewer ones. Border effect is often not very important among closely similar treatments. Hence the tendency has been to reduce plot size. Such factors as competition among plants, dust or spray drift, and insect migration, must be considered. It has been repeatedly shown that in field crop tests, long narrow plots are more efficient than square ones. Plots in critical experiments should not be made large simply to simulate commercial procedure, where small plots will answer the experimental questions. Large plots are often needed in extensive experiments, but in critical experiments, use of large plots often causes considerable loss in efficiency.

Several common mistakes in laying out experiments may be noted. The first is not considering the classification of experiments: putting treatments suited only for preliminary tests into the more mature critical experiments, or trying to make extensive experiments from critical ones. Lack of replication and randomization has already been discussed. Use of inadequate values for error is related to lack of replication. Occasionally in a large-scale test such as the forest treatment mentioned the experimenter may divide the large plots *after* treatment into a large number of small subplots. Taking results from these, considerable analysis is possible, but it can have no bearing on the principal result of the experiment. If treatment comparisons are based upon plots as units, error must also be based upon plots, not upon subplots.

If determination of plot values is based on sampling, an analysis of experimental error may be made just as in compound sampling (Chapter 2). The solution will help in determining the best number of plots per treatment and of samples per plot. Samples of this sort should be random if they are to be used in analysis of such error, but may be systematic if they are used merely to establish a plot mean. The same technique may be used to test effect of varying size of plots. If each plot is subdivided into several subplots, the analysis of error may be applied to determine the best combination of number of replications and number of subplots per plot.

3. Simple practical designs

The simplest design is that of unrestricted randomization, which corresponds to a completely random sample. In the field, with t treatments and n replications, the experimental area is divided into nt plots, and the several replications of each treatment are completely randomized among them. Several replications of one treatment may fall quite close together. The

analysis of variance of results is simply outlined as "between treatments" and "within treatments," the latter variance functioning as error. The design is used only occasionally in the field, since a moderate restriction on randomness usually results in a good gain in precision with lower error estimate. In the unrestricted random layout, the variation between parts of the area enters into error too much. On the other hand, it is flexible, and if several treatments have different numbers of replications, analysis is not hampered. It is often used in laboratory experiments. The insect measurements of Chapter 1 were compared in an unrestricted random scheme.

The second design to consider is the "randomized block," the most widely useful of all designs. It is an application of the principle of parallel comparisons. In the field, for t treatments and n replications, the area is divided into n blocks of t plots each. Each block is selected to be as uniform within as possible, and if there are some known differences in the area they should come between blocks. The t treatments are assigned randomly to the t plots of each block. The analysis of variance of results separates variance for blocks (uncontrolled field variation); for treatments (controlled), and their interaction. The last-named variance functions as error.

As a simple example, a test of 5 wheat varieties replicated 3 times may be cited. This calls for 3 blocks and 15 plots. The randomized field plan, with yields of varieties A to E, is given in table 1.

Analysis is shown in tables 2 and 3.

The F ratio for varieties is 22.6, highly significant for 4 and 8 d.f. The standard error of the difference between two variety means is $\sqrt{3.4\ (\frac{1}{3} + \frac{1}{3})}$ or about 1.5; multiplied by t for 8 d.f. (2.3), this gives us about 3.5 for least significant difference. Thus A and C are significantly above B and E, and D is below them all. Modification of least significant difference tests is noted in Chapters 1 and 4.

The blocks are designed to be as uniform within as possible, and to get most of the variation into the "between-blocks" category. If a known trend of variation crosses the area, the blocks may be laid off across the trend to get maximum difference, while the plots within each block may be laid off along the trend to make them as similar as possible. The gain is effi-

TABLE 1

Block I	Block II	Block III
B 20	C 28	A 33
D 18	A 30	E 26
A 28	E 23	B 28
C 29	D 16	C 30
E 20	B 26	D 19

TABLE 2. Variety yields collected

Variety ("treatment")	I	II	III	Total	Mean
A	28	30	33	91	30.3
B	20	26	28	74	24.7
C	29	28	30	87	29.0
D	18	16	19	53	17.7
E	20	23	26	69	23.0
Total	115	123	136	374	

The sum of squares of deviations: $(28)^2 + (20)^2 + \cdots + (26)^2 - (374)^2/15 = 9704 - 9325 = 379$.

Sum between blocks: $(115)^2/5 + (123)^2/5 + (136)^2/5 - (374)^2/15 = 45$

Sum between varieties: $(91)^2/3 + (74)^2/3 + \cdots + (69)^2/3 - (374)^2/15 = 307$

Sum for interaction, block × variety: $379 - 45 - 307 = 27$

TABLE 3. Summary of analysis of variance

Source of variation	Degrees of freedom	Sum of squares	Mean square
Blocks	2	45	22.5
Varieties	4	307	76.8
Interaction (Error in this case)	8	27	3.4

ciency in this design comes from testing treatments under as similar conditions as possible in each replication; this gain may be increased by a careful layout. It is common, with insect populations in the field, to find a directional trend in density which may be thus exploited.

In nonfield experiments where the uncontrolled variation may be subdivided thus, the randomized block plan can often be utilized. In animal feeding tests, the animals may be divided into groups according to gaining ability; the group then is the "block," the animal the "plot." In small animal tests the *litter* may act as the block; in skin tests where several are placed on an animal, one animal may be a block. Other cases might be cited; in laboratory tests a run or a day may be used as a block.

In insect toxicity tests the level of kill often varies from day to day; the chief cause appears to be variation as to resistance in the stocks available for testing. The days may be treated as blocks, with a full set of treatments each day in random order, and with essentially random drawing of insects from stock for test. Some old data on mosquito larvicidal tests are cited. Under present practice we would use several concentrations of larvicide and run a more elaborate experiment, but these data will illustrate the application (table 4).

TABLE 4. Percentage of mortality among mosquito larvae and its analysis

Insecticide ("treatment")	Day 1	Day 2	Day 3	Day 4	Day 5
A	27	40	12	57	30
B	58	82	48	88	55
C	58	68	33	80	67

	d.f.	S.S.	M.S.	F.
Between days	4	3322	830	24.4
Between insecticides	2	3164	1582	46.5
Day × insecticide (error)	8	276	34	

These percentage counts are based on adequate and similar numbers (100 each), and all are between 10 and 90%; so that analysis without transformation is all right. The insecticides were of course all at the same concentration and hence comparable. This illustrates use of the randomized block plan in a laboratory test.

This plan is very flexible. A whole treatment or replication may be omitted without hindering analysis. Snedecor (1956) gives instructions for analysis where one plot value is lost (a least square estimate of a value for carrying out symmetrical analysis), and for estimating the gain in efficiency over an unrestricted random experiment. The gain is usually a good one.

The Latin square is a design of greater restriction and narrower adaptation. The field plot layout for n treatments is a set of n^2 plots in n physical rows and n columns. The n treatments are assigned to n plots each, so that each treatment occurs once and only once in each row and each column, otherwise at random (a systematic rotation of assignment is undesirable). Table 5 gives a diagrammatic map of a typical Latin square.

In analysis of results, the total sum of squares of deviations is divided just as in table 2, but for rows, columns, and treatments; sum of squares for error is secured by subtraction of these from the total.

In spite of its glamorous name, the Latin square is adapted only to rather mature problems of 5 to 8 treatments. For these it tends to be a little more precise than the randomized block. It is harder to modify for omissions or losses, and is not good where unexpected upsets occur. Smaller squares

TABLE 5. Random Latin square for 5 treatments, A to E

A	B	C	D	E
C	E	A	B	D
B	C	D	E	A
D	A	E	C	B
E	D	B	A	C

have too few degrees of freedom for error; larger ones require too many replications.

The Latin square has only a limited use in nonfield experiments. In some cases, sources of variation corresponding to rows and columns can be recognized and removed, but interactions cannot be separated. In the field the rows and columns are merely subdivisions of field variation, so this "confounding" of interactions is not serious. In laboratory problems the confounding is often undesirable.

4. Complex practical designs

The split-plot design is one which always interests students, and has definite uses. In the field, the plots are divided into subplots, which are used for subsidiary experimental treatments, randomized among the subplots of each plot. Thus the plot for the major series of treatments becomes the block for the minor series, and there are two randomizations. Two classes of experiments are adapted to the plan. In one, some of the treatments require large units; tillage or irrigation are such treatments. At the same time other treatments which can be done in small units, such as fungicide or fertilizer, can be carried along. In the other type of tests, treatments which are known to differ widely can be estimated with a large error, while others can be studied more closely in subunits. Nonfield experiments of this type can be developed where conditions parallel those of the split-plot in the field. The splitting may go even farther than two stages (Goulden 1952).

As an example, suppose that in the wheat experiment of table 2, each of the 15 plots is divided into 3 parts, and 3 fungicidal treatments are randomized among the subplots of each plot. There will be then 45 yield values. The total sum of squares of deviations will be, for the 45, $S(X^2) - [S(X)]^2/45$, if the values are denoted by X's. There will be 15 plot values, each the sum of 3 X's; the sum of squares of deviations (S.S.) among them will be $[S(X)$ plot $1]^2/3 + [S(X)$ plot $2]^2/3 + \cdots + [S(X)$ plot $15]^2/3 - [S(X)]^2/45$. S.S. for blocks is $[S(X)$ Block I$]^2/15 + [S(X)$ Block II$]^2/15 + [S(X)$ Block III$]^2/15 - [S(X)]^2/45$. S.S. for varieties is $[S(X)$ variety A$]^2/9 + \cdots + [S(X)$ var. E$]^2/9 - [S(X)]^2/45$. The S.S. for interaction, block × variety, is: S.S. plots − S.S. varieties − S.S. blocks. This is "error A," the major plot error which is used to compare varieties.

Then a subsidiary table must be constructed of the 15 combinations of variety and fungicide, each the sum of 3 plots. For this 3 × 5 table with 15 "cells," S.S. for the table is $[S(X)$ cell$]^2/3 + \cdots + [S(X)$ cell $15]^2/3 - [S(X)]^2/45$. S.S. for fungicide treatments is $[S(X)$ fungicide $1]^2/15 + \cdots + [S(X)$ fungicide $3]^2/15 - [S(X)]^2/45$. The S.S. variety has already been

TABLE 6. Scheme of split-plot analysis

Source of variation	Degrees of freedom	Sum of squares, deviations as calculated	Mean square quotient, S.S./d.f.
Blocks	2	"	"
Varieties	4	"	"
B × V, Error A	8	"	"
Fungicides	2	"	"
F × V	8	"	"
Error B	20	"	"

calculated. The S.S. for interaction, variety × fungicide, is S.S. table − S.S. fungicide − S.S. variety. This is a subdivision of treatment effects, not an interaction used as error. The subplot error, or "error B," is a compound of interactions of "fungicide × block" and "fungicide × variety × block." These could be estimated by methods like those above, but it is convenient and quick to estimate Error B by subtracting all the S.S.'s already calculated from the total S.S.

The analysis may then be summarized (table 6). Fungicides and "F × V" are compared to Error B to test significance.

If actual calculations are made, they will be found to seem less complex than the outline above might appear. A worked example will be shown in Chapter 4.

The "switchback" type of experiment is adapted to some studies where readings are taken several times, especially in large animal experiments. Brandt's work on such problems is discussed by Cochran and Cox (1957). Treatments are rotated among experimental units. In entomology, in some tests of insect traps, rotation among trees and locations has been used. This is similar to the switchback. A type of analysis similar to the Latin square could be applied, accounting for variation among periods, locations, and trap types.

In some experiments such as fertilizer trials, several *levels* of application (such as none, single, or double) are often used. The sum of squares between levels may be divided to see (1) if there is a difference, (2) if the difference takes the form of a straight line, curve, etc. This obviously has important application to insecticide treatments, and leads toward study of dosage-mortality curves, which has been developed almost into a separate branch of biometrics.

The complex subject of analyses of combinations of similar experiments, conducted in different places or years, is well discussed by Cochran and Cox (1957). Experiments dealing with crop rotation (Crowther and Cochran

1942) are difficult to plan. There should if possible be enough plots to represent each phase of the rotation each year, and to provide replication.

Covariance analysis can add to the efficiency of some experiments, where some independent factor or "variate" is measured in each unit along with the variate of major interest. In stock feeding trials, the *initial weight* of each animal is sometimes taken to assist in the analysis of gain in weight. If initial weight has a definite correlation with gain, use of covariance will improve the comparison of gain.

A small experiment utilizing covariance analysis follows; other examples will be shown in Chapter 4. In laboratory spraying of a scale insect on twigs, living insects were counted on both sprayed and unsprayed portions of each twig. The circumstances were such that that sprayed area could not be counted before and after spraying; effectiveness was judged by comparing sprayed areas after spraying with unsprayed areas on the same twig. Three sprays were applied to 5 twigs each and numbers alive were counted. In practice, percentage of control would probably be estimated, but covariance is also a sound method of allowing for differences in original population. Tables 7 and 8 show data and analysis.

The analysis is simply set out as between treatments, and between twigs within treatments. The independent variate (X) is the number on unsprayed areas; the principal variate (Y) is the number alive on sprayed

Table 7. Live insects per unit of area

Treatment A		Treatment B		Treatment C	
Unsprayed	Sprayed	Unsprayed	Sprayed	Unsprayed	Sprayed
26	7	118	42	28	2
26	4	24	2	48	5
86	54	78	15	88	32
52	17	72	0	24	5
86	10	42	3	28	11
—	—	—	—	—	—
276	92	334	62	216	55

Grand totals: 826, 209.

Table 8. Summary of covariance analysis

Source of variation	Degrees of freedom	$S(x^2)$	$S(xy)$	$S(y^2)$	Errors of estimate, $S(y^2) - [S(xy)]^2/S(x^2)$		
					d.f.	S.S.	M.S.
Sprays	2	1393	+87	155	—	—	—
Within sprays (error)	12	11,694	+4918	3504	11	1436	131.0
Total	14	13,087	+5005	3659	13	1745	—
Difference for test	—	—	—	—	2	309	154.5

areas after spraying. The total S.S. for X is $(26)^2 + \cdots + (28)^2 - (826)^2/15 = 13{,}087$. S.S. between sprays is $(276)^2/5 + (334)^2/5 + (216)^2/5 - (826)^2/15 = 1393$. The sum within sprays is $13{,}087 - 1393 = 11{,}694$. In the same way the sums of squares for Y are calculated as 3659, 155, and 3504. Analysis of square roots might be a little more precise, but does not seem necessary in this illustrative example.

Next we need the sums of *products* of deviations of X and Y (S.P. or $S(xy)$). The total S.P. is secured as in regression in Chapter 1; $(26 \times 7) + (26 \times 4) + \cdots + (28 \times 11) - (209 \times 826)/15$ which is + 5005. The sum of products between sprays is $(276 \times 92)/5 + (334 \times 62)/5 + (216 \times 55)/5 - (826 \times 209)/15$, or + 87. The S.P. within sprays is 5005 − 87 or + 4918. Thus the S.P. is partitioned.

The operations are carried on as indicated in the table. One degree of freedom is subtracted under "Errors of estimate" because of allowance for regression as well as for mean.

Without the covariance, the mean squares for *sprays* and error would be 155/2 or 77.5, and 3504/12 or 292.0 respectively; with the covariance, these values are 154.5 and 131.0. Although the difference is not significant in either case, the greater tendency to significance and the lower error show what may be expected of covariance where there is a real correlation. Note that total or "treatment + error" and "error" are the only classifications used in adjustment. This is true also in more complex analyses; blocks, or rows and columns, are omitted from "errors of estimate" analysis as extraneous.

Covariance is not itself an experimental design, except that the plan must include measurement of the independent variable in each unit. It can be used with various designs, and will often add information and avoid the necessity of a more complex plan. In the above example the adjustment was for simple linear regression. More complex analyses sometimes adjust for a curved relation or for more than one independent variable. Treatment means may be adjusted, as shown by Snedecor (1956), to the values expected when x is at the mean.

5. *Factorial design and confounding*

The idea of testing all combinations of 2 or more factors is basic in factorial design. The simplest combination is 2×2, as when 2 concentrations of an insecticide are tested, each with and without a supplement. The "$2 \times 2 \times 2$" fertilizer trial is a classic for study; all combinations of nitrogen, phosphorus, and potassium are tried at 2 levels. The levels may be merely "none" and "some." If manure, present or absent, is included in the test, we have a $2 \times 2 \times 2 \times 2$, or "$2^4$." Often there will be more than 2 levels; $2 \times 2 \times 3$, $3 \times 3 \times 3$, or other combinations.

These plans are of value because they give estimates of interactions or differential effects. They must of course be replicated in some good design such as the randomized block. When several quantitative levels of a factor are used, there is a possibility of estimating trend, both linear and curvilinear.

The technique of analysis known as the use of "single degrees of freedom" has a special application to such problems. Snedecor (1956, Ch. 12, or Ch. 15 in earlier editions) describes it. Where no confounding is present, a sum of squares can be derived for each individual degree of freedom; and if combinations are properly chosen, the sums will add to the correct total. For the randomized blocks problem of table 2, with block totals of 115, 123, and 136, there are two degrees of freedom for blocks. Each block has 5 values. The first degree of freedom is chosen as comparison of Block I with II; the second to compare Blocks I and II with Block III (table 9).

The two sums of squares are each calculated as (net sum)^2/divisor. They total 44.9, while the rounded sum for blocks in table 3 is 45.

The divisors are the sums of squares of the number of times each block is used (coefficients), times the number of items in each. For the second degree of freedom, the divisor is $[(-1)^2 + (-1)^2 + (+2)^2]\ 5$, or 30. The net sum is $-115 - 123 + 2(136)$ or $+34$.

The comparisons must be chosen by certain rules. There are $(n - 1)$ total d.f. Use any element by itself but once; it may be later used in combinations, but only once in the same combination. All must be used somewhere. In the block analysis, Blocks I and II are used by themselves in the first degree of freedom. In the second, they are combined against III. The coefficients in any row must add to zero; the products of corresponding coefficients in any two rows must add to zero. For example, in table 9, products of corresponding coefficients in the one below with the one above, are $(-1) \times (+1), (-1) \times (-1), (+2) \times (0)$; or $-1, +1, 0$. Ones and zeros are often omitted from the table, ones appearing as simply $+$ or $-$.

The treatment sum of squares of an insecticide-supplement 2×2 experiment may be illustrated. The degrees are set up to give effect of concentration, of supplement, and of interaction (table 10). The coefficients for the interaction may be secured by multiplication of the corresponding coefficients directly above them.

The single-degree method is occasionally useful to work out some special

TABLE 9. Single-degree analysis of blocks

Block I (115)	Block II (123)	Block III (136)	Net sum	Divisor	(Net sum)^2/divisor
+1	−1	0	−8	10	$8^2/10 = 6.4$
−1	−1	+2	+34	30	38.5

TABLE 10. Scheme of single-degree analysis, 2 × 2 factorial

	C_1	C_2	$C_1 + S$	$C_2 + S$
1st degree—concentration	−1	+1	−1	+1
2nd degree—supplement	−1	−1	+1	+1
3rd degree—interaction, (C × S)	+1	−1	−1	+1

comparison. It is also valuable in working out new experimental schemes. A randomized block experiment can be completely analyzed thus, with every degree of freedom for treatment, error, etc. In the Latin square the error degrees cannot be separated because of confounding.

Many other and more complex factorial plans and analyses could be described. They can be studied in sources given, especially Yates (1937) and Cochran and Cox (1957). It is possible to analyze factorials to bring out many interesting relations. In general, as complexity grows with greater numbers of factors and levels, there is less freedom in choice of combinations. The analysis must follow "formal" subdivisions as given in instructions.

In complex experiments the number of treatments may grow quite large. If in the field, many plots per block are needed; the large blocks grow variable and part of the advantage of the block plan is lost. It is usually found that high-order interactions (triple, quadruple, etc.) are not higher in variance than error, and are of small importance. By carefully looking over these combinations some degrees of freedom of probable small importance may be selected to sacrifice by "confounding." Thus the total number of plots needed, and block size, is reduced, and the important comparisons are saved.

A simple illustration will be shown using the classic N, P, K fertilizer test, with each element at two levels (table 11). Coefficients for main effects are of course derived by writing −1 for the low level, +1 for the high. For interaction, they are products of two corresponding coefficients; for example,

TABLE 11. Single treatment degrees of freedom, 2 × 2 × 2

d.f.	$N_1P_1K_1$	$N_2P_1K_1$	$N_1P_2K_1$	$N_1P_1K_2$	$N_2P_2K_1$	$N_2P_1K_2$	$N_1P_2K_2$	$N_2P_2K_2$
N effect	−1	+1	−1	−1	+1	+1	−1	+1
P effect	−1	−1	+1	−1	+1	−1	+1	+1
K effect	−1	−1	−1	+1	−1	+1	+1	+1
N × P	+1	−1	−1	+1	+1	−1	−1	+1
N × K	+1	−1	+1	−1	−1	+1	−1	+1
P × K	+1	+1	−1	−1	−1	−1	+1	+1
N × P × K	−1	+1	+1	+1	−1	−1	−1	+1

TABLE 12. Analysis of 2 × 2 × 2, confounded and unconfounded (4 complete replications)

Unconfounded,	d.f.	Confounded,	d.f.
Blocks	3	Blocks	3
Treatments	7	Subblocks within blocks	4
Error	21	Treatments	6
		Error (residual)	18

N × P in the first column is (−1) × (−1) or (+1). Coefficients for N × P × K are the products of those for N and P × K, or P and N × K, etc.

To sacrifice the triple interaction, the 8 plots or units of each replication block are divided into halves of 4 units each. The N × P × K is estimated (table 10) by contrasting 2 groups: ($N_2P_2K_2$, $N_2P_1K_1$, $N_1P_2K_1$, $N_1P_1K_2$) vs. ($N_1P_1K_1$, $N_2P_2K_1$, $N_2P_1K_2$, $N_1P_2K_2$). If in each complete replication one of these groups is randomized in one half-block, the other in the other half-block; the triple interaction is completely mixed up or *confounded* with "half-blocks within blocks." If there are real differences between the subblocks, a gain in precision will result from a lower error estimate. Table 12 shows outline of analysis of variance each way.

Many confounded designs of this sort have been worked out, some of them very complex in plan and analysis. They have not been very widely used, and apparently are used more in laboratory than in field tests. There are numerous examples of factorial experiments in entomology, but not as a rule any involving confounding. The example discussed in the introduction to this work was a 2 × 2 factorial.

Of late years modified factorials have been developed to estimate optimum combinations of treatments by multiple regression techniques; this is the study of "response surfaces." An extreme application of factorials called "fractional replications" has been used, estimating main effects and some important interactions using other interactions as an error estimate. This seems better adapted to engineering experiments than to biology.

6. Incomplete block designs

Study of confounded designs has given rise to a special group of designs called by Cochran and Cox (1957) "incomplete block designs," although all confounded plans feature imcomplete blocks in some way. The designs we speak of here are adapted to experiments with a large number of treatments, not necessarily nor usually factorial in arrangement. They were originally designed to compare large numbers of newly developed crop varieties, keeping blocks small. In the usual analysis the small blocks totals are adjusted for treatments, and after analysis the treatment means are in turn adjusted for blocks. In many of the designs the blocks of an entire rep-

lication can be kept together; and if the blocks within replications do not vary significantly, the experiment can be analyzed as a randomized block.

If the small blocks do vary significantly, the error variance will be lowered as compared with the randomized block analysis.

Adjustments of the sort mentioned can be made (or attempted) in ordinary experiments which have become mixed up, but this analysis is a miserable business. With the planned incomplete block designs, a number of special short-cut methods make analysis smoother. These methods will not work for all numbers of treatments, and the plan must be carefully worked out and adhered to. These designs have more risks of failure than simple designs, and their use will hardly pay unless the number of experimental treatments exceeds 15. With large numbers of treatments they offer a chance of gain. Cochran and Cox (1957) and Fisher and Yates (1963) list a great many such plans and their analyses in detail. With large numbers of treatments, it is usually possible to add or omit borderline cases to make up a desired number.

One type is called the "balanced incomplete block" plan. Every *pair* of treatments occur together the same number of times. This condition is called "balance." Making up sets reminds one of parlor puzzles, but the principal workable combinations have been tabled. After sets are made up, treatments are randomized in blocks. This design is sometimes used in laboratory trials, with a day or run constituting the block, as well as in the field. In some balanced incomplete block plans, complete replications can be kept together; in others they cannot. A very simple example follows, of 6 treatments, a to f, 3 in a block, with 5 replications: abc, abd, ace, adf, aef, bcf, bde, bef, cde, cdf.

The *lattices* are so called because the groups of sets cross each other. The typical lattice has k^2 treatments, k in each block. The sets can be made up from a $k \times k$ Latin square written down on paper. In the simplest case one group of sets for the small blocks is made from rows, the other from columns. A third group can be made from *letters* of the Latin square, (the A's, B's, etc.) to form a *triple* lattice. Carrying the process further, groups can be made from other criteria till all possible pairs are together—a *balanced* lattice. An orthogonalized Latin square (Fisher and Yates 1963) is useful in working out a balanced lattice. In the ordinary simple or triple lattice all pairs do not occur together, but analysis is not very difficult. In lattices, after the sets are made up, the treatments are randomized in each small block, and the blocks in the replications. Replications can be kept together. Number of treatments is usually a perfect square, though rectangular lattices are sometimes used.

In *lattice squares*, the field plots are laid out in actual rows and columns in the field. The small blocks constitute rows or columns, or sometimes rows

and columns. Balance is achieved, complete replications are kept together, and directional variation is removed as in the Latin square. Lattice squares can be made up from orthogonalized squares, but the principal plans are ready made in Cochran and Cox's text (1957). The 6 × 6 and 10 × 10 will not work for lattice squares or balanced lattices.

There are numerous variants of these and other incomplete block designs described by Cochran and Cox and other writers (more are being added each year), but these three seem the most likely to be widely useful. Some of them have been used in entomological field plot trials with some gain in accuracy (Wadley 1945 and 1946). A couple of examples of analysis of such experiments are given in Chapter 4.

7. Literature cited

Years ago, long lists of references would have been needed in such a discussion as this, but the appearance of Cochran and Cox's textbook (1950, 2nd ed. 1957) has simplified the task of citing literature. Their account is very complete, with instructions for layout and analysis of many plans, and references to original literature. Several good texts on experimental design have appeared lately, but for the beginner in these concepts Cochran and Cox's book seems unequaled. A few foundation articles (Yates 1933 and 1937), a few statistical texts of experimental slant (Fisher 1956, Goulden 1952, Snedecor 1956, Fisher and Yates 1963) and some articles on special situations (Crowther and Cochran 1942, Wadley 1945 and 1946) are cited. Fisher's "Design" (1960) with its philosophy must be included.

Cochran, W. G., and Cox, Gertrude M.
1957. Experimental designs. 2nd ed., 611 pp. Wiley, New York.
Crowther, F., and Cochran, W. G.
1942. Rotation experiments with cotton in the Sudan Gezira. Jour. Agr. Sci. (England) 32: 390–405.
Fisher, R. A.
1956. Statistical methods for research workers. 12th ed., 356 pp. Oliver and Boyd, London.
1960. The design of experiments. 7th ed. 248 pp. Oliver and Boyd, London.
Fisher, R. A. and Yates, F.
1963. Statistical tables for biological, agricultural, and medical research. 6th ed., 146 pp. Hafner, New York.
Goulden, C. H.
1952. Methods of statistical analysis. 2nd ed., 467 pp. Wiley, New York.
Snedecor, G. W.
1956. Statistical methods. 5th ed., 534 pp. Iowa State, Iowa.
Wadley, F. M.
1945. Incomplete block experimental designs in insect population problems. Jour. Econ. Ent. 38: 651-654.

1946. Incomplete block design adapted to paired tests of mosquito repellents. Biometrics Bull. 2: 30-31.

Yates, F.

1933. The principles of orthogonality and confounding in replicated experiments. Jour. Agr. Sci. (England) 23: 108–145.

1937. The design and analysis of factorial experiments. Imper. Bur. Soil Sci. (England), Tech. Commun. 35.

CHAPTER 4

Practical Examples

This chapter consists of a series of practical problems which have actually arisen, worked out in detail from start to finish. It is hoped that (with previous examples) they will illustrate a variety of situations and will help in approaching new problems.

1. Binomial distribution and mortality in successive samples

A problem occurred in work with Japanese beetle larvae in which 40 groups of 5 insects each were treated by a certain method. Results were as follows:

Sets with all 5 killed	8
" " 4 killed 1 alive	10
" " 3 " 2 "	9
" " 2 " 3 "	7
" " 1 " 4 "	5
" " all 5 surviving	1

The total mortality was 126 out of 200 or 63 percent. What would be the expectancy of survival distribution with only random variation among a uniform lot of insects? If the population *is* uniform with 63 percent susceptible, the only reason for variation is that more susceptible insects will be drawn in some samples of 5 than in others. This corresponds to the variation observed in shaking pennies and counting heads, and is the lowest variation we can expect. The binomial $(p + q)^n$ is here $(0.63 + 0.37)^5$. Expanding, we write $(0.63)^5 + 5(0.63)^4(0.37) + 10(0.63)^3(0.37)^2 + 10(0.63)^2(0.37)^3 + 5(0.63)(0.37)^4 + (0.37)^5$. The first term gives the expectation of proportion all dead, no survivors; the second term, the proportion with 4 dead and 1 alive; etc. Calculation gives these terms as 0.099, 0.291, 0.342, 0.201, 0.059, 0.007 respectively; they total 0.999, or about 1, as they should. Multiplying each by 40, the number of sets; we get 4.0, 11.6, 13.7, 8.0, 2.4, and 0.3 as the average number out of 40 expected in each class.

Lastly we may test by chi-square whether our actual distribution could

TABLE 1

Number dead	Frequency observed (O)	Frequency calculated (C)	O − C	(O − C)²	(O − C)²/C
5	8	4.0	+4.0	16.00	4.00
4	10	11.6	−1.6	2.56	0.22
3	9	13.7	−4.7	22.09	1.61
2	7	8.0	−1.0	1.00	0.12
1	5	2.4	+3.3	10.89	4.03
0	1	0.3			
Sum	40	40.0	0.0	—	9.98

TABLE 2

p	F	Fp	p^2	Fp^2
100	8	800	10,000	80,000
80	10	800	6,400	64,000
60	9	540	3,600	32,400
40	7	280	1,600	11,200
20	5	100	400	2,000
0	1	0	0	0
Total	40	2520	—	189,600

be derived by random sampling from a population with the distribution given (table 1).

Chi-square is 9.98; there are 5 final classes, and two degrees of freedom are lost, for p and for total number. A value of 9.98 with 3 d.f. is significant, but not highly so; the 5 percent point is 7.82. We might possibly secure a distribution like this from a uniform population of 63 percent susceptibility, but the probability of doing so is not very great. It is more likely that there were some differences among the assemblage of grubs sampled, and thus differences appeared among the sets greater than could be explained by random sampling.

This could be treated by computing variance among the 40 samples and comparing this variance to the theoretical variance. With percentage counts this theoretical variance may be expressed as pq/n; here it is $(63 \times 37)/5$, or 466, if stated in percentage form. The standard deviation corresponding is $\sqrt{466}$ or about 21.6. We have 8 cases with all 5 dead (100 percent mortality), 10 cases with 4 dead or 80 percent, etc. (table 2).

Sum of squares of deviations is $189{,}600 - (2520)^2/40$, or 30,840. Variance is 30,840/39 or 791. This is significant as compared to 466 in the "F" test with 39 and ∞ degrees of freedom.

2. Meaning of zero secured in sampling proportions

Suppose a group of 80 insects gives zero survival when treated with an insecticide. How high a real survival might conceivably be present in the population? To test the possibility of 2 percent, the binomial is $(0.98 + 0.02)^{80}$, and only the first term need be calculated. This is $(0.98)^{80}$, and is the expected proportion of times when a sample of 80 from a population, of which 2 percent are resistant, would give *no* survival. It can be solved by logarithms; the answer will be the antilogarithm of $[80 \times \log (0.98)]$. It comes out to be 0.199 or nearly 20 percent. Of samples of 80 from a population 2 percent resistant, about one-fifth will have no survivors. Hence a single sample with 100 percent mortality is not strong evidence against average survival as high as 2 percent.

Let us try 5 percent. The term $(0.95)^{80}$ is calculated as 0.016. Less than 2 percent of samples of 80 from a population 5 percent resistant would be expected to give zero survival. Hence getting a complete kill among 80 is pretty good evidence against average survival as high as 5 percent.

The meaning which can be ascribed to zero in a limited sample may thus be worked out by trial and error. However, a general statement may be made that with a real survival of 3 individuals or less, zero may often occur in a single sample. With a real survival of 4 or more, zero is unusual. This will hold for samples of various sizes and varying percentage levels. It is associated with the similarity of the binomial and Poisson series at low percentage values.

The problem may be looked at another way in planning experiments. Numbers may be provided which will give zero survival, if secured, a definite meaning. With only 30 insects in a test, for example, we will occasionally expect a complete kill, when the real population value is 3 survivors or 10 percent. Hence with only 30, zero survival will be fairly good evidence against over 10 percent survival, but will not be strong evidence against survival as high as 10 percent. With 300 insects, zero survival by the same standard will be evidence against a true survival above 1 percent. Chitwood and Blanton (1941) have applied this relation to the problem of numbers necessary in tests.

In making such applications it should be remembered that expected distributions will apply only when all individuals are strictly from the same population. We might try 300 insects in a single sample, and from zero survival we might conclude that true survival was 1 percent or less in the population. That would indeed be true of the exact population furnishing the sample, but not necessarily for all insects of the species. On repeating the test the next week, we might run into a population of somewhat greater resistance. For broad conclusions about a problem it is necessary to have

replicated experiments with repeated sampling, exposing experimental treatments to the kind of variation they will face in practice.

Tests of the kind outlined will be useful in figuring the minimum numbers that could give a satisfactory answer.

3. Poisson series; worms per 100 apples

If an insect population is distributed fully at random, with equal and constant probabilities for each unit, numbers per unit will be in a Poisson distribution.

One hundred apples sampled from a large population showed worm populations as in table 3. The Poisson frequency series is calculated as follows: for 0, $e^{-\bar{x}}$; for 1, $\bar{x}e^{-\bar{x}}$; for 2, $\bar{x}^2e^{-\bar{x}}/2$, etc. In general it is $\bar{x}e^{-\bar{x}}/X!$; where X is the number per unit, $X!$ is factorial X, and $\bar{x}$ is the mean. The expression $e^{-\bar{x}}$ is equal to $1/e^{\bar{x}}$; "e" being about 2.72, the base of natural logarithms. The solutions are proportions or fractions of the total.

In our problem, $\bar{x} = 0.31$; $e^{\bar{x}}$ is computed as $2.72^{0.31}$, the antilogarithm of $(0.31 \times \log 2.72)$; this is the antilogarithm of (0.31×0.4343) or of (0.1346), which is about 1.363. This gives $e^{\bar{x}}$ as 1.363; $e^{-\bar{x}}$ would be 1/1.363 or 0.734. We expect a zero in 0.734 or 73.4 percent of the cases if the material shows a true Poisson distribution. Multiplying 0.734 by the total $n(100)$ we get the expected number or frequency as 73.4 compared to the actual number of 81.

The remaining frequencies can be readily calculated from the zero frequency. The expected proportion of apples with one worm is $(\bar{x}e^{-\bar{x}})/1$, or 0.31×0.734. The expected proportion of twos is $(\bar{x}^2e^{-\bar{x}})/(1 \times 2)$; of threes, $(\bar{x}^3e^{-\bar{x}})/(1 \times 2 \times 3)$; etc. So each term can be gotten from the preceding one by multiplying the numerator by $\bar{x}$ and the denominator by X, the number of which the frequency is sought. When the proportion is secured,

TABLE 3

No. worms (X)	No. apples (F)	FX
0	81	0
1	12	12
2	4	8
3	2	6
4	0	0
5	1	5
	100	31

$S(X) = 31$

$\bar{x} = S(X)/n = 31/100 = 0.31$

it may be multiplied by the total number of apples to turn it to an expected frequency (table 4).

The distribution actually found may be compared with the Poisson expectation for this mean and total by use of chi-square (table 5). Enough "lumping" of higher classes is carried out to give an expectation well over 1 in the smallest class.

Chi-square is $S[(O - C)^2/C]$ or 8.37. With the "lumping" used, we have 3 classes remaining. Two degrees of freedom are lost, because the fitted distribution must agree with the actual in *mean* and in total *number*. This leaves 1 d.f. A chi-square of 8.37 with 1 d.f. is highly significant, well beyond the 1 percent point (see table in Chapter 1). This tells us that the distribution of worms we got, in the sample of 100 apples, could hardly have been sampled from material in the Poisson distribution with a mean of 0.31 worm per apple. The inference is that the population is not Poisson in distribution; probabilities are not constant for all apples.

The difference is seen to be in the existence of more zeros and more high frequencies in the actual population than the Poisson (pure random distribution) would give. This is typical of insect population counts. A simple way to compare an actual distribution with the Poisson is to calculate the variance of the actual distribution. It may be compared with the expected Poisson variance, which equals the mean. This is done using the "F" test with $(n - 1)$ degrees of freedom for the actual variance, and an infinite number for the Poisson variance. In the example given, $S(X^2)$ is $(81 \times 0) + (12 \times 1^2) + (4 \times 2^2) + (2 \times 3^2) + (1 \times 5^2)$, or 71; $[S(X)]^2/n = (31)^2/100$ or 9.61. $S(x^2)$ is 71 − 9.61 or 61.39; $V = 61.39/99$ or 0.62. The F calculated is 0.62/0.31 or 2.00, as the mean is 0.31. This F is highly significant with 99 and ∞ degrees of freedom. Chi-square can be adapted to this same test of homogeneity as noted by Snedecor (1956).

4. Theoretical minimum Poisson variance as a tentative substitute for computed sampling variance in a population estimate

To compute sampling variance it is necessary to have some sort of repeated sampling from the same material. In estimating insect population

Table 4

No. worms X	Numerator $e^{-\bar{x}}\bar{x}^x$	Denominator $X!$	Proportion	Frequency (100 × Proportion)
1	$0.734 \times 0.31 = 0.2275$	1	0.2275	22.8−
2	$0.734 \times (0.31)^2 = 0.0705$	1×2	0.0352	3.5
3	$0.734 \times (0.31)^3 = 0.0219$	$1 \times 2 \times 3$	0.0036	0.4−
4	$0.734 \times (0.31)^4 = 0.0068$	$1 \times 2 \times 3 \times 4$	practically 0	0.0+
above 4			practically 0	

TABLE 5

Worms per apple	Frequency of apples		O − C	$(O - C)^2$	$(O - C)^2/C$
	Observed (O)	Calculated (C)			
0	81	73.4	+7.6	57.76	0.79
1	12	22.8	−10.8	116.64	5.12
2 & up	7	3.9	+3.1	9.61	2.46
Total	100	100.1	—	—	8.37

for an area, repeated composite samples, all from the same material, furnish an excellent basis for computing variance. Single sample units may be used if selected by a random or restricted random plan (Chapter 2). However, in some cases we have only a single composite sample, not planned with the idea of computing sampling error, and a tentative approximation may help.

Cotton gin trash inspections for the presence of pink bollworm larvae were examined in several Texas counties. Inspection results in one year are shown in table 6.

In thinking of sampling error we think of the reproducibility of the sample if taken repeatedly from the same material. Each of these samples was taken only once. If successive thoroughly distributed samples were taken their totals would tend to fall in something like the Poisson series. The variance in this series is equal to the mean. In County B, for example, a series of samples of 1103 bushels each would be expected to have a mean somewhere near 6 larvae per sample, a variance in the series near 6, and a standard deviation somewhere near $\sqrt{6}$ or about 2.5.

If we state the number of larvae as 6 ± 2.5, we can carry through the calculation of larvae per bushel as about 0.005 ± 0.002 for County B. Treated in this way the other density estimates are stated as 0.781 ± 0.043 for A, 0.004 ± 0.002 for C, and 0.001 ± 0.0004 for D. The significance of the difference of A from all the others is plain; B and C are substantially equal. It seems likely that the difference of D from B and C may be due only to sampling error.

County A in successive years gave results as in table 7.

Using the same procedure, there is no doubt of a real difference between year 1 and 2, also between 1 and 3. No difference can be detected between 2 and 3.

This procedure should not be substituted for computed sampling error when the latter can be provided, but will often give a useful tentative idea when computed sampling error is not available. It may be added that the ability to compute sampling error depends on planning of the sampling.

TABLE 6

County	Trash inspected (bushels)	Larvae found (number)	Larvae per bushel
A	415	324	0.781
B	1103	6	0.005
C	1161	5	0.004
D	5403	5	0.001

TABLE 7

Year	Bushels examined	Larvae found	Larvae per bushel
1	415	324	0.781
2	1109	5235	4.720
3	928	4408	4.750

5. The Poisson series in a problem of parasite distribution

A problem often discussed in insect ecology is the degree of selectiveness possessed by an adult parasite. Can it avoid hosts already attacked, thus preventing overlapping and waste of eggs? Some evidence was secured by a forest insect worker studying an injurious caterpillar attacked by a fly parasite. The egg-shells of the parasite could be found on the caterpillar's skin, thus giving a record of attempts to parasitize each host. The record was not perfect, as shown by an occasional parasitized host without egg-shells, but was evidently fairly accurate. It may be examined for what it can show.

If there is perfect discrimination on the part of the parasite adult, no host will receive two eggs as long as any have none. If there is no special discrimination, all hosts having an equal probability of being attacked, the distribution of eggs will be close to the Poisson series. If some larvae have higher probability of attack than others, there will be more overlapping than the Poisson series would indicate. This last condition is what we usually meet in insect population sampling in the field.

The number of egg-shells on host larvae are given in table 8. The total is 236 eggs on 555 larvae, or a mean ($\bar{x}$) of 0.425. The proportion having zero in a Poisson series would be estimated as $e^{-\bar{x}}$, which is $(2.718)^{-0.425}$, or $1/(2.718)^{0.425}$. This can be solved logarithmically as 1/antilog (0.4343×0.425), which is 0.6536. The expected number of larvae with zero is the proportion 0.6536 times the total number 555, or 362.7.

The proportion having one egg is estimated as $e^{-\bar{x}}\bar{x}$, or 0.425×0.6536, which is 0.2778. The expected number with one is 555×0.2778, or 154.2.

In the same way the proportions with two ($e^{-\bar{x}}(\bar{x})^2/2$), with three ($e^{-\bar{x}}(\bar{x})^3/3!$), and so on, are calculated. Multiplication of proportions by the total number gives the expected number (table 9). Lumping the small expectancies, 4 classes result. Chi-square is calculated with two degrees of freedom, and is quite significant.

There is a tendency to more zeros (omissions) and more high values (duplications) than pure random distribution would give. The parasite is apparently not able to discriminate much, and its egg-distribution is a little worse than a random one.

6. Meaning of zero in population samples

If we can define the level of a population which may be regarded as unimportant, we can plan sampling so that it will be likely to detect any important population, as stated in a previous chapter. *Well-distributed* sampling is not likely to give a zero where the true value is 5. If populations above 0.01 larva per bushel (see example 4) must be detected without much chance of failure, 500 bushels should be taken. If this number gives zero, the level is probably below 0.01 per bushel. To be reasonably sure of detecting infestations as low as 0.001 per bushel, 5,000 bushels would be needed.

TABLE 8

Number of egg-shells	Number of host larvae
0	390
1	120
2	29
3	9
4	4
5	3

TABLE 9

Number of eggs	Number of host larvae			
	Observed		Calculated	
0	390		362.7	
1	120		154.2	
2	29		32.8	
3	9	16	4.6	5.2
4	4		0.5	
5 and up	3		0.1	

7. Negative binomial of Fisher fitted to wormy-apple series

This distribution is described as $(q - p)^{-k}$ where p and k are computed from mean and variance ($\bar{x}$ and V), and $q = 1 + p$. In this series (example 3) $\bar{x} = 0.31$; $V = 0.62$; $p = (V - \bar{x})/\bar{x} = (0.62 - 0.31)/0.31 = 1.000$; $q = 2.000$; $k = (\bar{x})^2/(V - \bar{x}) = 0.0961/0.31 = 0.310$. The probability of each number, P_x, is computed as $q^{-k}\,[(k + X - 1)!/X!(k - 1)!]\,(p/q)^x$.

For X of zero, since $0! = 1$, this whole expression cancels out to q^{-k}, or $1/q^k$, which is $1/(2.000)^{0.31}$. The expression $(2.000)^{0.31}$ = antilog of $0.31 \times$ log 2.000 or of 0.301×0.31, which is antilog 0.09331, or 1.24. $1/(2.000)^{0.31}$ is 1/1.24 or 0.806.

For P_x when $X = 1$, we have $0.806 \times (0.310!)/(1!)\ (-0.69!) \times (1.000/2.000)$. It is necessary to understand fractional factorials to solve this, but it comes out as 0.125.

For P_x when $X = 2$, the expression is $0.806 \times (1.310!)/(2!)\ (-0.69!) \times (1.000/2.000)^2$, which equals 0.04. Higher P's will be small and can be lumped; subtracting P_0, P_1 and P_2 from 1.000, we have P for 3 and higher as 0.028.

Assembling our P's and multiplying by N, we get the estimated frequency (table 10). In computing chi-square, we need 4 final classes to have one degree of freedom, since the expected and actual are made to agree in mean, variance, and total number. Chi-square will be small and nonsignificant for 1 d.f.

This distribution seems to show good agreement with the actual numbers in such insect populations.

8. Comparison of mortality at several temperatures

A large population of tropical insects was given a severe treatment consisting of prolonged exposure to constant low temperatures. The aim was to secure 100 percent mortality. Several temperatures from 32° to 35°F. were used, and samples withdrawn at several time intervals from 1 to 10 days. With the earlier withdrawals, some difference in mortality between temperatures was seen, with higher mortality at lower temperatures. As the point of complete mortality was approached, the difference between

TABLE 10

X (No. of worms)	P	$P \times N$ = F expected	F actual
0	0.806	80.6	81
1	0.125	12.5	12
2	0.041	4.1	4
3 and up	0.028	2.8	3

TABLE 11

Temperature	Total number insects	Number surviving (X)	Percent surviving (p)	pX
32°	9091	9	0.099	0.891
33°	5542	11	0.198	2.178
34½°	1375	3	0.218	0.654
35°	5337	9	0.169	1.521
Total	21,345	32	—	5.244

TABLE 12

Temperature	O (actual)	C (expected)	O − C	$(O - C)^2/C$
32°	9	13.6	−4.6	1.56
	9082	9077.4	+4.6	<0.01
33°	11	8.3	+2.7	0.88
	5531	5533.7	−2.7	<0.01
34½°	3	2.1	+0.9	0.39
	1372	1372.9	−0.9	<0.01
35°	9	8.0	+1.0	0.12
	5328	5329.0	−1.0	<0.01
		Chi-square		2.95+

temperatures seemed to be equalized. The results at 8 days were tested by chi-square using the procedure of Snedecor's text (1956) discussed in Chapter 1. $\bar{p}$ is derived as 32/21,345, or 0.150 percent. Chi-square is

$$\frac{100[S(pX) - \bar{p}S(X)]}{\bar{p}(100 - \bar{p})} = \frac{100[(5.244) - (0.150 \times 32)]}{0.150 \times 99.850}$$

$= 44.4/14.98 = 2.96$ with 3 d.f. This is not significant and indicates that such numbers surviving might all have been secured from the same population in several samples. In other words, no difference in results between temperature in this range is shown.

This problem might have been approached in other ways. The expected survival numbers might be computed for each temperature, multiplying $\bar{p}$, that is, 0.150 percent, by the total number used. Chi-square could then be calculated as in previous examples, from "O" and "C" numbers. The binomial standard deviations (pq/n) could be calculated for the percentages and would tend to show that the percentages were not really different.

Below is shown the same problem using O and C numbers. In calculation of this type expected and actual numbers must be used for both *surviving* and *dying;* though in this particular case, use of numbers dying makes little

difference. Expected numbers are calculated from average mortality, 0.15% (table 12).

9. Simple variation

A problem in sampling population density of aquatic insect larvae was presented a few years ago. The population was fairly well distributed in the area studied but the work of sampling and counting was very heavy. It was desired, in preliminary work, to bring the standard error of the mean to a level of 10 percent of the mean.

Fourteen standard dredge counts were made, well distributed through the field of inquiry. The readings were used in calculating the variance and standard deviation as in Chapter 1 (table 13). The desired standard error of 10 percent of the mean would be 10.8. In the equation $s_{\bar{x}} = s/\sqrt{n}$, we may write the $s_{\bar{x}}$ desired (10.8) and the s estimated from the sample (26.7), and solve for n:

$$10.8 = 26.7/\sqrt{n}; \quad 10.8\sqrt{n} = 26.7;$$

$$\sqrt{n} = 26.7/10.8 = 2.47; \quad n = 2.47^2 = 6.10$$

An n of 6 will come near the desired $s_{\bar{x}}$ and one of 7 should give a lower one.

TABLE 13

Sample number	Larvae, number (X)	X^2
1	118	13,924
2	70	4,900
3	99	9,801
4	124	15,376
5	129	16,641
6	96	9,216
7	86	7,396
8	94	8,836
9	100	10,000
10	134	17,956
11	89	7,921
12	140	19,600
13	162	26,244
14	75	5,625
Total	1516	173,436

Mean, $\bar{x} = S(X)/n = 1516/14 = 108.3$.

Sum of squares of deviations from mean, $S_{(x^2)} = S(X^2) - [S(X)]^2/n = 173{,}436 - 164{,}161 = 9{,}275$.

Variance, $V = S_{(x^2)}/(n-1) = 9{,}275/13 = 713.46$.

Standard deviation, $s = \sqrt{V} = \sqrt{713.46} = 26.7$.

The application here is not only to the specific sample and to the tests of significance which may be made, but to the reproducibility of future samples from the material. The estimated standard error gives the scope of expected variation of future samples from the true mean.

Nearly all samples will be expected to fall within about 2 standard errors of the true mean. The multiplier is given by the "t" table (Chapter 1); for 13 degrees of freedom it is about 2.2. We use 13 d.f. here because our estimate of the standard deviation is no better and no worse than that provided by 14 cases. With 6 units, and an estimated standard error of 10.9, we will expect nearly all to fall within about 24 points above or below the true mean. These are the so-called confidence limits.

The expected standard error of a difference of two means of n cases each is $\sqrt{2V/n}$ or $\sqrt{2}$ times the standard error of a mean. The difference which should be recognized as significant in such material will be about twice the standard error of a difference. Multiplying $s_{\bar{d}}$, or 15.3, by 2.2, we get 34, the least difference between two simple means of 6 cases each which will be recognized as significant. It will be seen that this is about 3 times $s_{\bar{x}}$, or in this case 30 percent of the mean.

Cochran (see Ladell 1938) has extended this line of thought to give an estimate of the largest real difference in the population likely to be missed by sampling. It is somewhat larger than the least significant difference.

10. Compound population sampling

The simple case of example 9 is much less frequent in population studies than more complex cases. In regional surveys, fields are our usual units and the sampling within fields is of subunits. The error affecting the area mean must be estimated from variation between fields, not by that within fields.

An example is drawn from grasshopper egg-sampling data on a species with rather wide field distribution of eggs. Ten fields with 5 square-foot units in each were studied (table 14).

The variation within fields has an *influence* on that between fields; $s_{\bar{x}}$ for the area mean would be smaller if fields were determined perfectly instead of by sampling. In this analysis the calculated $s_{\bar{x}}$ includes some variation due to within-field variation. Later the estimation of the share of variance from within will be used in problems. It often points the way to more efficient work, but cannot change the verdict of work already done. Increasing the within-field sampling in the case above would make field means more stable, and would tend to decrease $s_{\bar{x}}$; but $s_{\bar{x}}$ would still be based on between-field variation.

These statistics are the sort for which the transformation $\sqrt{n + 0.5}$ is

TABLE 14

Field number	Subunits, e.g. Number egg-pods					Field mean
1	0	0	1	0	0	0.2
2	3	0	1	2	1	1.4
3	2	0	1	0	0	0.6
4	0	0	0	0	0	0.0
5	0	1	0	0	0	0.2
6	0	2	2	0	0	0.8
7	0	1	0	0	5	1.2
8	1	1	0	0	0	0.4
9	2	7	0	0	1	2.0
10	0	1	0	2	1	0.8

TABLE 15

Source of variation	Degrees of freedom	Mean square original counts	Mean square $\sqrt{n + 0.5}$
Between fields	9	1.95	0.2485
Within fields	40	1.79	0.2122
F	—	1.09	1.17

usually recommended, so that 0 becomes 0.707, 1 becomes 1.225, etc. However, we are estimating only the area mean and its error, and the total egg-pods for this purpose number 38. Hence the average should be rather stable even without transformation.

For field totals:

$$S(x^2) = 1^2 + 7^2 + \cdots + 4^2 - 382/10 = 87.60$$

$$V\bar{x} = 87.60/(9 \times 10) = 0.9733;\ s\bar{x} = 0.986$$

$$t = 2.26.\ \text{Confidence limits} = \pm 2.26 \times 0.986 = 2.23\ (9\ \text{d.f.})$$

$$\text{mean} = 3.80\ \text{with C.L. } 6.03\ \text{and } 1.57$$

On square-foot basis ($\div 5$) this is 0.76 with C.L. 1.21 and 0.31.

Using the square root transformation $n + 0.5$, this calculation becomes $1.225^2 + 2.734^2 +$ etc. The mean and confidence limits are calculated, then squared, and from each is subtracted 0.5, arriving at the original scale. When put on a square-foot basis as before, the mean is estimated as 0.65 with C.L. 0.28 and 1.14. Hence transformation produced little change in conclusions.

Analysis of variance of individual values turns out as shown in table 15. The verdict is similar by the two methods. Note that S.S. between fields in original counts (17.52) is $\frac{1}{5}$ the S.S. found for field sums above (87.60).

11. Standard error of a difference

Elytron length was measured on two groups of beetles believed to be specifically distinct, though similar (table 16). It was desired to test the difference; and if it were real, to find diagnostic characters.

The mean difference is 31.04 − 27.30, or 3.74. The pooled variance is (61.3036 + 34.475)/(6 + 9) or 6.3852. The variance of the difference is 6.3852 ($\frac{1}{7}$ + $\frac{1}{10}$) or 1.5510; $s_{\bar{d}}$ is $\sqrt{1.5510}$ or 1.25. Since t is 3.74/1.25 significance is reached with 15 d.f. Although the difference is significant, the measurement would not be at all useful for a diagnostic character because of overlapping of individual values.

This problem will be discussed further under Section 32.

12. A simple correlation problem

Examples of correlation studies have already been touched upon in Chapter 1. A correlation analysis was recently employed in a case involving the relation of results with a rapid method of estimation to results with a more laborious and exact method. Visual examination of cornstalks attacked by corn borer was compared with a method of dissecting stalks and counting larvae. The visual estimate is taken as the independent variable, since the plan is to estimate population by this easier method, and is denominated as X. The dissection count becomes Y.

TABLE 16

	Group A	Group B
Individual measurements (in micrometer units)	26.75	25.25
	28.25	25.50
	28.75	26.00
	31.00	26.50
	33.75	27.00
	34.00	27.00
	34.75	27.00
		28.25
		28.50
		32.00
Sum, $S(X)$	217.25	273.00
n	7	10
$\bar{x}$	31.04	27.30
$S(X^2)$	6803.8125	7487.3750
$S(X)^2/n$	6742.5089	7452.9000
$S(x^2)$	61.3036	34.4750

TABLE 17. Number of larvae

By visual examination	Dissected	By visual examination	Dissected
18	11	9	6
9	11	7	4
8	6	2	2
10	6	15	13
17	12	7	11
35	24	9	17
11	8	2	1
3	3	7	7
2	2	5	7
11	15	8	5
8	8	5	4
29	26	5	4
6	5		

For one series of small plots numbers of larvae are given in table 17. Derived statistics are as follows:

$$S(X) = 248; \bar{x} = 9.92$$

$$S(X^2) = 3960$$

$$S(x^2) = 3960 - (248)^2/25 = 1500+$$

$$S(Y) = 218; \bar{y} = 8.72$$

$$S(Y^2) = 2892$$

$$S(y^2) = 991$$

$$S(XY) = 3240$$

$$S(xy) = 3240 - (218 \times 248)/25 = +1077$$

Plotting as in Chapter 1 shows a rather definite but not exact relation with no sign of departure from linearity. The swarm of dots seems to spread out more widely with high values. The correlation coefficient, r_{xy}, $= S(xy)/\sqrt{S(x^2)S(y^2)} = +1077/\sqrt{991 \times 1500} = +0.88$. This is highly significant with 23 degrees of freedom and shows a real and positive relation as might be expected. The regression coefficient $S(xy)/S(x^2)$, is $+1077/1500$ or $+0.718$. The regression equation is $Y - \bar{y} = b(X - \bar{x})$; since $b = 0.718$, $\bar{y} = 8.72$, and $\bar{x} = 9.92$, we simplify this to $Y = +1.60 + 0.718X$.

The sum of squares of error of estimate of Y is calculated as $S(y^2) -$

$[S(xy)]^2/S(x^2)$, or 991 − 773, or 218. The variance of estimate is this sum divided by $n - 2$, or 9.48; the standard error of estimate ($s_{y.x}$) is $\sqrt{9.48}$ or about 3.1. The standard error of the regression coefficient is $s_{y.x}/\sqrt{S(x^2)}$ or $3.1/\sqrt{1500}$, which is 0.080. The regression coefficient, $+0.718 \pm 0.080$, therefore is significantly lower than 1.

In applying the correlation study to the objective of the problem, we must watch for (a) bias and (b) inexactness. If the laborious dissection and counting is taken as a standard of exactness, it is evident that on the average the visual estimate is a little too high; the regression coefficient is significantly less than one. The estimate even allowing for this is not very exact, with a standard error for individual plot estimates of about 3 larvae. From the plotted points we may deduce that estimates are more accurate with low populations. The scatter tends to be wider at higher populations. Thus the material does not quite fulfill the requirements for an accurate correlation study. While visual estimates are usually higher than the counts, they fall considerably lower in a few cases, showing signs of being a little erratic at times.

The relation is a real one and can be employed in a limited range of populations, with due allowance for any bias, and some sacrifice of accuracy. The saving in work may make possible the examination of more material to offset the latter's sacrifice.

13. A multiple correlation problem

Data on soil surface temperature were published a few years ago, with corresponding air temperatures and depth of snow. A part of the data are shown in table 18. The relation of both snow depth and air temperature to soil surface temperature may be investigated by multiple correlation methods.

First, sums of squares of deviations and of products of deviations are calculated as in the preceding example. However, this time there are 6 sums as follows:

$$S(y^2) = 59.0$$

$$S(x_1^2) = 6922$$

$$S(x_2^2) = 257.3$$

$$S(x_1y) = 493.2$$

$$S(x_2y) = -59.3$$

$$S(x_1x_2) = -702.3$$

TABLE 18

Temperature of soil, °F. (Y)	Temperature of air, °F. (X_1)	Snow depth inches (X_2)
28	+8	8
30	+10	9
30	+8	10
30	+16	9
30	+22	9
30	+32	8
31	+36	7
31	+38	6
32	+36	6
32	+38	5
32	+40	3
32	+47	0
33	+52	0
34	+47	0
33	+40	0
32	+30	2
31	+29	1
30	+20	2
28	+19	2
28	+11	2
31	+23	4
30	+8	7
30	−10	6
29	−13	6

The estimating equation will be $Y = a + b_1X_1 + b_2X_2$. The two net or partial regression coefficients, or "b's," are calculated from the following equations:

$$S(x_1^2)b_1 + S(x_1x_2)b_2 = S(x_1y)$$

$$S(x_1x_2)b_1 + S(x_2^2)b_2 = S(x_2y)$$

Substituting the values calculated, we have

$$6922b_1 + (-702.3b_2) = +493.2$$

$$-702.3b_1 + 257.3b_2 = -59.3$$

Solving, $b_1 = +0.066$, $b_2 = -0.050$

The "a" is derived from the b's and means; $a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2$, or $a = 30.71 - [(+0.066) \times 24.46] - [(-0.050) \times 4.67]$. This is 29.3, and the equation is $Y = 29.3 + 0.066X_1 - 0.050X_2$.

For any given values of air temperature and snow depth (X_1 and X_2) the expected Y (soil temperature) may be calculated. If we calculate a value for Y for each of the 24 cases in the table, we can compare actual and expected Y and calculate a sum of squares of deviations of actual from expected. This can be done by a short-cut process, however, by an extension of methods of simple correlation. The part of the total $S(y^2)$ accounted for by the relation is estimated as $b_1S(x_1y) + b_2S(x_2y)$, or 35.5. Subtracting this from $S(y^2)$, we get 59.0 − 35.5 = 23.5. This is the part of variation not accounted for by the relation. If estimated Y is denoted by y_e, the standard error of estimate is $\sqrt{S[(Y - y_e)^2]/(n - 3)}$, since 3 constants are calculated; it is $\sqrt{23.5/21}$, or a little over 1.0.

The multiple correlation coefficient R can easily be estimated. The part of variation accounted for by the relation is $R^2S(y^2)$, and R^2 = (part accounted for)/$S(y^2)$. Here we have R^2 = 35.5/59.0 or 0.602, and R = $\sqrt{0.602}$ or 0.776. Its significance can be determined by analysis of variance using part accounted for (2 d.f.), or by tables in Snedecor's older editions. In this case it is quite significant.

The simple correlation coefficients can easily be determined from sums computed already; r_{y1} = +0.77, r_{y2} = −0.48, and r_{12} = −0.53. It will be noted that r_{y1} is nearly as high as our multiple correlation coefficient $R_{y\cdot12}$. This shows that no great gain was made by taking snow cover into the calculations. This may be tested by analysis of variance, testing variance accounted for against remaining variance. Sum of squared deviations from simple regression of Y on X_1, can be calculated as in the preceding example (table 19). Mean square C is tested by the "F" table against mean square B. It is highly significant. Mean square E is tested against D. It is not significant.

Thus it is seen that snow cover in these particular data is not important and did not add to accuracy of estimation of soil temperature.

Snedecor (1956) gives these and also more complex methods, which are

TABLE 19

Source of variation	Degrees of freedom	Sum of squares	Mean square
A. Total, $S(y^2)$	23	59.0	—
B. Deviations from simple regression Y on X_1	22	23.9	1.09
C. Simple regression (A − B)	1	35.1	35.10**
D. Deviations from multiple regression	21	23.5	1.12
E. Gain from multiple regression (B − D)	1	0.4	0.40

** Highly significant.

very workmanlike and lead to all the above determinations. They also show the calculation of "betas" or standard partial regression coefficients. Many will prefer the more elaborate methods, but the ones shown are sufficient.

It is not difficult to extend these methods to more than two independent variables. For three, for instance, equations are

$$S(x_1^2)b_1 + S(x_1x_2)b_2 + S(x_1x_3)b_3 = S(x_1y)$$

$$S(x_1x_2)b_1 + S(x_2^2)b_2 + S(x_2x_3)b_3 = S(x_2y)$$

$$S(x_1x_3)b_1 + S(x_2x_3)b_2 + S(x_3^2)b_3 = S(x_3y).$$

The *simple* and *multiple* correlation coefficients have been computed. The *partial* correlation coefficients may also be mentioned. The simple correlation coefficient of soil temperature and air temperature disregards snow cover. Their partial coefficient gives their estimated correlation if snow cover were held constant. Where we have three simple correlation coefficients, (r_{12} , r_{13} , r_{23}), we may estimate the partial correlation of 1 and 2, for example, by a short-cut formula. It is

$$r_{12.3} = (r_{12} - r_{13}r_{23})/\sqrt{(1 - r_{13}^2)\,(1 - r_{23}^2)}$$

For soil and air temperatures, we calculate

$$r_{y1.2} = (+0.77 - (-0.48)(-0.53))/$$

$$\sqrt{[1 - (-0.53^2)][1 - (-0.48^2)]} = +0.69$$

For soil temperature and snow cover, $r_{y2.1} = -0.13$, not significant.

The apparent negative correlation of snow cover and soil temperature is due only to the correlation of soil and air temperatures, in connection with the correlation of snow depth and air temperature. The relation of soil and air temperatures, on the other hand, is real and definite, with or without allowance for effect of snow cover.

14. Analysis of variance, simple classification, irregular class numbers

Insect mortality counts in laboratory experiments on scale insects were expressed in percentages. Since percentages were based on adequate and similar numbers (about 100), and since practically all were between 10 and 90 percent, analysis of untransformed figures was used. Results were as follows:

Treatment A: 15 %, 12 %, 8 %, 21 %	total 56, mean 13.5
Treatment B: 27 %, 33 %, 36 %, 41 %, 39 %	total 176, mean 35.2

Treatment C: 63 %, 54 %, 88 % total 205, mean 68.3
Grand total: 437
Total sum of squares: $(15)^2 + (12)^2 + \cdots (88)^2 - (437)^2/12$
$= 21{,}819 - 15{,}914$ (about) $= 5905$
Sum of squares between classes: $(56)^2/4 + (176)^2/5$
$+ (205)^2/3 - 15{,}914$
$= 784 + 6195 + 14{,}008 - 15{,}914 = 5073$
Sum of squares within classes $= 5905 - 5073 = 832$

Summary

	Degrees of freedom	Sum of squares	Mean square
Between classes	2	5073	2536+
Within classes	9 (3 + 4 + 2)	832	92+

The "F" is 2536/92, highly significant with 2 and 9 degrees of freedom. This will serve to illustrate this type of analysis. A chi-square test would not be adequate here, since classes are not internally homogenous in susceptibility. The analysis above correctly points out significant differences between methods.

15. Analysis of variance, random block experiment, compared to "Student's method" of pair differences

It has been stated that analysis of variance of cross-classified material is an extension of the use of variance of a series of differences. It can be seen that if only two things are compared either method could be used, but that the pair difference method will not work for more than two at a time.

In an experiment on cotton insects, two methods of control were used side by side in each of 10 fields. The fields can be regarded as blocks, the treatment areas as plots. Results are shown in table 20. The variance of pair differences can be calculated from the "difference" column. The sum of squares of deviations is $(+59)^2 + (+40)^2 + \cdots + (+56)^2 - (359)^2/10$, which is 2811. The variance of the series of differences is 2811/9 or 312+; the variance of the mean difference is 312/10 or 31.2. The standard error is $\sqrt{31.2}$ or 5.6. The "t" value is $\bar{d}/s_{\bar{d}}$, 35.9/5.6 or 6.41; compared with a value necessary for significance of 2.26 with 9 d.f.

The analysis of variance proceeds in the standard manner. The total sum of squares of deviation is $(68)^2 + (60)^2 + \cdots + (28)^2 - (885)^2/20$, or 9146. The sum between blocks is $[(77)^2 + (80)^2 + \cdots + (112)^2]/2 - (885)^2/20$, or 1297. The sum between methods is $[(622)^2 + (263)^2]/10 -$

TABLE 20

Block No.	Percent of squares attacked			Difference (A − B)
	Method A	Method B	Sum	
1	68	9	77	+59
2	60	20	80	+40
3	76	24	100	+52
4	45	28	73	+17
5	59	28	87	+31
6	38	18	56	+20
7	66	32	98	+34
8	54	48	102	+6
9	72	28	100	+44
10	84	28	112	+56
Total	622	263	885	359
Average	62.2	26.3	—	35.9

TABLE 21

Source of variation	Degrees of freedom	Sum of squares	Mean square
Total	19	9146	—
Blocks	9	1297	144+
Methods	1	6444	6444
Error (Interaction, block × method)	9	1405	156+

$(885)^2/20$, or 6444. The sum for interaction (error) is derived by subtraction as 9146 − 1297 − 6444, or 1405 (table 21).

"F" for methods is 6444/156, or 41.3, highly significant with 1 and 9 d.f.; the 5% point (table 10, Chapter 1) is only 5.1 in the table.

Note that our calculated F is equal to t^2; 6.41 squared is equal to 41.3 within the allowance of small errors caused by rounding. The tabular F is also equal to the corresponding t^2. If the standard error of a difference is determined from the analysis of variance, we have $\sqrt{156(1/10 + 1/10)}$ or $\sqrt{31.2}$ as before. Note that 156 is one-half the 312 of the other method but has a smaller divisor. It may also be noted that the division into blocks did not add to the precision of the test in this particular experiment, since they did not vary more than random error would indicate.

This shows the relation and equivalence of the two methods. Analysis of variance has of course much broader possibilities than the pairing method.

This and the previous example illustrate the cases of cross-classified and unordered ("nested") classes respectively. Note differences in computation of sums of squares.

A more complex case is one in which each subclass has more than one determination. Suppose that in the 2 × 10 test just discussed, each value is based on 2 samples (40 in all). Then after calculating the total sum of squares of deviations among the 40, each of the 20 cells would be summed. A sum of squares for this table would be calculated; (Total for subclass $1)^2/2 + \cdots +$ (total for subclass $20)^2/2$ − correction factor. Then the total S.S. for methods (2 methods with 20 cases each) and for blocks (10 of 4 cases each) would be calculated. The interaction S.S. is S.S. table − S.S. blocks − S.S. methods.

Then subtracting S.S. table from S.S. total, we have left S.S. deviations between items within subclasses. It could be determined independently from subclass means, but is easier to get by subtraction.

The use of these special tables to calculate interaction S.S. will be developed further in Examples 17 and 18. We have cases where both unordered and cross-classified situations occur in the same complex analysis (Cassil, Wadley, and Dean 1943). In the case discussed there, we had both plots and trees within plots, interacting with *section* of the trees (top, middle, or bottom). The S.S. for *trees within plots* was calculated by getting S.S. for *all trees*, and subtracting from this S.S. for *plots*. The interaction is likewise handled; S.S. interaction (all trees × sections) − S.S. interaction (plots × sections) yields S.S. interaction (trees within plots × sections).

16. Analysis of variance in a compound classification of plots, without restriction to blocks

The experiment in question was on scale insect populations on fruit trees. It was rather exceptional in that it was compound, with several trees per plot, and a number of leaf sample records for each tree, and that block restrictions were not used. For this reason no interactions can be determined. We have variance between treatments, between plots within treatments, trees within plots, and leaf-units within trees. This makes the analysis an interesting exercise in methods. A more efficient plan was recommended to the experimenters, and in later work single-tree plots arranged in blocks were used.

We will consider 5 treatments, with 4 plots per treatment, 2 trees per plot, and a sample of 20 leaves per tree. Counts were mostly of a high level not needing transformation for analysis. For brevity, the individual leaf records are not given; but only the sums and sums of squares, of population counts, for each treatment. There are in all 800 items. The 2 tree totals are shown for each plot, followed by their sum in parenthesis, and separated from the next plot by a semicolon (table 22). The sum, and the sum of squares of all individual leaf records, are shown in the last two columns.

The total sum of squares of deviations can be derived from the totals of these last two columns, as: $3,611,423 - (33,107)^2/800$, or 2,241,331. This is in fact the sum of squares of the 800 leaf records, and includes the other sources of variation.

Next comes the sum of squares for treatments: $[(3956)^2 + (4633)^2 + \cdots + (12,391)^2]/160 - (33,107)^2/800$, or 293,819. (There are of course 160 units in each treatment, giving the divisor.) The sum of squares for all 20 plots is: $[(834)^2 + (705)^2 + \cdots + (2221)^2]/40 - (33,107)^2/800$, or 588,413. This, however, *includes* the sum of squares for treatments. The sum for *plots within treatments* is 588,413 − 293.819, or 294,594.

The sum of squares for all 40 trees is $[(553)^2 + (281)^2 + \cdots + (892)^2]/20 - (33,107)^2/800 = 686,659$. This includes sums for plots and treatments, so the sum for *trees within plots* is 686,659 − 294,594 − 293,819, or 98,246. The sum of squares for leaf units *within trees* will be given by subtracting the 3 sums already derived, from the total. It comes out as 1,554,672.

The summary is in table 23.

The variance for plots within treatments is the appropriate error for treatments, since it includes all the uncontrolled or sampling variation. A treatment based on plot units must have an error in plot units. Variation of trees may be regarded as error for plots, and leaves as error for trees. The latter variances are interesting and useful in study of technique, but are not adequate as error for treatments. A test will show plots as significantly higher in variance than trees, and tree variance as significantly higher than leaf variance.

TABLE 22

Treatment	Totals for 2 trees in each of 4 plots	$S(X)$	$S(X^2)$
"O"	553, 281 (834); 404, 301 (705); 609, 756 (1365); 215, 837 (1052)	3,956	212,652
"L"	510, 898 (1408); 482, 193 (675); 1544, 746 (2290); 106, 154 (260)	4,633	374,083
"U"	567, 521 (1088); 262, 82 (344); 628, 701 (1329); 1258, 1034 (2292)	5,053	334,753
"W"	1493, 1326 (2819); 557, 354 (911); 389, 265 (654); 1545, 1145 (2690)	7,074	596,176
"T"	1241, 1504 (2745); 2971, 1645 (4616); 1175, 1634 (2809); 1329, 892 (2221)	12,391	2,093,759
Total		33,107	3,611,423

TABLE 23

Source of variation	Degrees of freedom	Sum of squares	Mean square
Total	799	2,241,331	—
Treatments	4	293,819	73,455*
Plots within treatments	15	294,594	19,640
Trees within plots	20	98,246	4,912
Leaves within trees	760	1,554,672	2,046

* Significant, F = 73,455/19,640.

17. Analysis of variance, 4 criteria of classification

An experiment in insect winter survival was run during 4 years, with 3 depths of burial, 3 moisture conditions, and 2 dates of burial. Only one unit was provided for each combination. The uncontrolled variation measured was that due to years, and interaction with years furnished error variances for several questions. These interactions included whatever sampling variation occurred in insect populations, site selection, etc.

The figures were in percentage survival. Each percentage was based on several hundred insects. Since percentages were often low (0 to 3 %) and occasionally as high as 15 or 20 %, the angle transformation was used. The 72 percentages of survival when transformed are shown in table 24. First the sum of squares of deviations among all 72 is computed as $S(X^2) - [S(X)]^2N$ or 17,511.46 − $(883.8)^2/72$ which is 17,511.46 − 10,848.64 or 6,662.82. Next 6 two-way groupings must be arranged for the 6 two-way interactions: Date-number (of irrigations), date-depth, date-year, number-depth, number-year, depth-year. Each entry in table 25 is the sum of 12 items. For December with 2 winter irrigations, for example, we add the 12 items for this combination in 4 years and at 3 depths: 5.3 + 14.1 + ··· 9.4 + ··· 4.4 + ··· + 4.0, giving us 74.2.

Analysis of the table is as follows:

Total S.S.:

$$[(258.1)^2 + (114.9)^2 + \cdots + (106.0)^2]/12 - 10{,}848.64 = 2107.72$$

S.S. between dates:

$$(447.2)^2/36 + (436.6)^2/36 - 10{,}848.64 = 1.56$$

S.S. between numbers:

$$(469.6)^2/24 + (234.0)^2/24 + (180.2)^2/24 - 10{,}848.64 = 1974.37$$

S.S. interaction, date × number: 2107.72 − 1.56 − 1974.37 = 131.79

TABLE 24

Burial date	Winter irrigations (number)	Depth (inches)	1936	1937	1938	1939
Dec.	2	2	5.3	14.1	18.0	0.0
		4	9.4	4.1	8.1	2.3
		6	4.4	1.5	3.0	4.0
	1	2	14.6	22.1	16.6	7.3
		4	3.3	16.1	13.4	5.2
		6	2.9	4.6	6.5	2.3
	0	2	24.9	40.7	36.2	31.7
		4	23.9	26.9	19.2	10.9
		6	13.9	16.2	6.7	6.9
Jan.	2	2	10.4	25.2	17.2	9.9
		4	4.8	9.0	12.8	0.0
		6	5.8	4.8	0.0	6.1
	1	2	14.1	27.4	28.4	16.2
		4	8.6	6.6	6.5	5.2
		6	3.8	0.0	0.0	2.3
	0	2	23.8	36.9	22.2	26.7
		4	18.9	13.1	12.1	10.1
		6	11.5	14.0	7.8	14.4

TABLE 25. Date-number

Date	0	1	2	Total
Dec.	258.1	114.9	74.2	447.2
Jan.	211.5	119.1	106.0	436.6
Total	469.6	234.0	180.2	883.8

Note that 12 items make up each cell total, 36 make up each date total and 24 each number total, and that these are used as denominators. This keeps the analysis in terms of original units, and we use the same correction factor throughout.

We then proceed to arrange the other five groupings (table 26). There is a good deal of overlapping and opportunity for checking. For the date-depth grouping 12 items are added for each cell, for all years, and numbers of winter irrigations.

To get the 4 triple interactions it is necessary to have 4 tables; each lumping one factor and classified according to the other three. After getting the

TABLE 26

DATE-DEPTH

Date	2 inches	4 inches	6 inches	Total
Dec.	231.5	142.8	72.9	447.2
Jan.	258.4	107.7	70.5	436.6
Total	489.9	250.5	143.4	883.8

Total S.S.	= 2704.58
S.S., dates	= 1.56
S.S., depths	= 2622.85
S.S. interaction, date × depth, 2704.58 − 1.56 − 2622.85	= 80.17

NUMBER-YEAR

Number	1936	1937	1938	1939	Total
0	116.9	147.8	104.2	100.7	469.6
1	47.3	76.8	71.4	38.5	234.0
2	40.1	58.7	59.1	22.3	180.2
Total	204.3	283.3	234.7	161.5	883.8

Total S.S.	= 2529.38
S.S., numbers	= 1974.37
S.S., years	= 438.23
S.S. interaction, number × year (by subtraction)	= 116.78

DEPTH-YEAR

Inches	1936	1937	1938	1939	Total
2	93.1	166.4	138.6	91.8	489.9
4	68.9	75.8	72.1	33.7	250.5
6	42.3	41.1	24.0	36.0	143.4
Total	204.3	283.3	234.7	161.5	883.8

Total S.S.	= 3513.23
S.S., years	= 438.23
S.S., depths	= 2622.85
S.S. interaction, depth × year	= 452.15

DATE-YEAR

Date	1936	1937	1938	1939	Total
Dec.	102.6	146.3	127.7	70.6	447.2
Jan.	101.7	137.0	107.0	90.9	436.6
Total	204.3	283.3	234.7	161.5	883.8

Total S.S.	= 489.78
S.S., years	= 438.23
S.S., dates	= 1.56
S.S. interaction, date × year	= 49.99

TABLE 26—*Continued*

NUMBER-DEPTH

Number	2 inches	4 inches	6 inches	Total
0	243.1	135.1	91.4	469.6
1	146.7	64.9	22.4	234.0
2	100.1	50.5	29.6	180.2
Total	489.9	250.5	143.4	883.8

Total S.S.	= 4824.44
S.S., numbers	= 1974.37
S.S., depths	= 2622.85
S.S. interaction, number × depth	= 227.22

TABLE 27. Date-number-year

Date	Number	1936	1937	1938	1939	Total
Dec.	0	62.7	83.8	62.1	49.5	258.1
	1	20.8	42.8	36.5	14.8	114.9
	2	19.1	19.7	29.1	6.3	74.2
Jan.	0	54.2	64.0	42.1	51.2	211.5
	1	26.5	34.0	34.9	23.7	119.1
	2	21.0	39.0	30.0	16.0	106.0
Total		204.3	283.3	234.7	161.5	883.8

sum of squares among the items of the table, the sums of squares for the main effects and two-way interactions involved are deducted. The remainder is the triple interaction.

In table 27, depths are added for each subclass, so that 3 items are added to give each cell total. In the first item (Dec. burial, 1936, no winter irrigation) the addition is 24.9 + 23.9 + 13.9 = 62.7. Sum of squares for the table is $[(62.7)^2 + (83.8)^2 + \cdots + (16.0)^2]/3 - 10{,}848.64$; or 2784.36. Subtract from this the sums of squares already secured for main effects, date, number, year; and for interactions date × number, date × year, number × year. The remainder is for triple interaction, date × number × year; 2784.36 − 1.56 − 1974.37 − 438.23 − 131.79 − 49.99 − 116.78 = 71.64. (Note that the bottom totals check with others for years in the first series of tables; the side totals with the cells of one of the two-way tables.)

Tables for date-year-depth, number-year-depth, and date-depth-number may be arranged in the same way (table 28). Lastly, the sum of squares for quadruple interaction is secured by subtracting from the total sum of squares of deviations first determined (6662.82) the 14 sums of squares for

TABLE 28

DATE-DEPTH-YEAR

Date	Depth (inches)	1936	1937	1938	1939
Dec.	2	44.8	76.9	70.8	39.0
	4	36.6	47.1	40.7	18.4
	6	21.2	22.3	16.2	13.2
Jan.	2	48.3	89.5	67.8	52.8
	4	32.3	28.7	31.4	15.3
	6	21.1	18.8	7.8	22.8

Total S.S. = 3679.66 and S.S. for triple interaction = 34.71 in the same way as before.

NUMBER-DEPTH-YEAR

Number	Depth (inches)	1936	1937	1938	1939
0	2	48.7	77.6	58.4	58.4
	4	42.8	40.0	31.3	21.0
	6	25.4	30.2	14.5	21.3
1	2	28.7	49.5	45.0	23.5
	4	11.9	22.7	19.9	10.4
	6	6.7	4.6	6.5	4.6
2	2	15.7	39.3	35.2	9.9
	4	14.2	13.1	20.9	2.3
	6	10.2	6.3	3.0	10.1

Total S.S. = 5965.40; triple interaction S.S. = 133.80

DATE-NUMBER-DEPTH

Date	Number	2 inches	4 inches	6 inches
Dec.	0	133.5	80.9	43.7
	1	60.6	38.0	16.3
	2	37.4	23.9	12.9
Jan.	0	109.6	54.2	47.7
	1	86.1	26.9	6.1
	2	62.7	26.6	16.7

Total S.S. = 5179.37; triple interaction S.S. = 141.41

4 main effects, 6 two-way interactions, 4 triple interactions. It comes out as 186.15.

These calculations are summed up in table 29. (Pooled triple and quadruple interactions, mean square 14.2, 40 d.f.) It is believed that this summary shows the tendencies well, and in these extreme percentages we may have

TABLE 29

Source of variation	Degrees of freedom	Sum of squares	Mean square
Between dates	1	1.56	1.56
Years	3	438.23	146.07**
Depths	2	2622.85	311.42**
Numbers	2	1974.37	987.18**
Date × year	3	49.99	16.66
Date × depth	2	80.17	40.08*
Date × number	2	131.79	65.90*
Year × depth	6	452.15	75.36*
Year × number	6	116.78	19.46
Depth × number	4	227.22	56.81*
Date × year × depth	6	34.71	5.78
Date × number × depth	4	141.41	35.37
Date × year × number	6	71.64	11.94
Depth × year × number	12	133.80	11.15
Depth × date × year × number	12	186.15	15.51
	—	———	
Total	71	6662.82	—

* Significant. ** Highly significant.

somewhat more confidence in relations as shown by the transformed function than by the original figures.

In evaluating significance, the principle of selecting error terms expressing the random variation will apply. Depths, numbers of irrigations, and dates are reproducible experimental treatments, while years are nonreproducible, and bring in uncontrolled sampling variation. The proper error for depths, numbers, and dates respectively is their interaction with years, since this expresses the variation we would have to contend with in doing the work over; i.e., the error for "depths" is "depth × year," etc. The proper error for years is the variance for triple or quadruple interaction, since every other main factor may be reproduced, and two-way interactions involving years are with reproducible sources. It is accordingly somewhat easier to prove that years are really different, with the other sources of variation exactly reproducible, than to say that treatments are really different, when they have to be compared in variable years.

The two-way interactions involving only depth, number, and date are referred to the corresponding triple interactions with years as error. Those involving years are referred to quadruple interaction. They express differential effect. In date × depth, for example, we not only had a strong in-

fluence of depth, but depth tended to act in a different way with different dates. (Note that date is lower than its error but not significantly so; "F" is 16.66/1.56 with only 1 d.f. for lesser mean square, 3 for greater.) Other significant interactions can be similarly interpreted.

Triple interactions are referred to quadruple interaction as error; none proves significant. In short-cut procedure all could be pooled with the quadruple as error for two-way interactions. If a triple interaction was significant, it would mean that a two-way interaction varied with the level of the third factor.

Standard errors for treatment means will have only a limited value, since significant interactions with other treatments occur. A given treatment mean will tend to vary if other treatments are shifted, and our evaluation of them applies only to the present set-up. However, a standard error and confidence limits will be calculated to illustrate the procedure with transformations.

For 0 irrigation, the mean survival (degrees) is 19.6;
The variance applying is year × number, 19.46 (6 d.f.);
The variance of the mean is 19.46/24 for 24 items, or 0.817;
The standard error is $\sqrt{0.817}$ or 0.904;
t for 6 d.f. is 2.45 (5% level);
95% confidence limits are $\pm 2.45 \times 0.904$, or 2.2;
or from 21.8 to 17.4.

Going to Snedecor's equivalent angle table, we have:

	Degrees	%
mean	19.6	11.3
upper C.L.	21.8	13.8
lower C.L.	17.4	8.9

18. A split plot experiment in insect population

A South American experiment in cotton insect population is available for study. Effect of irrigation water, fertilization, and spacing was studied. The criterion was seasonal average of percentage of forms attacked by bollworm. The nature of the irrigation treatment made it necessary to have large plots, while small plots were sufficient for the other treatments. There were 3 levels of water application (W_1, W_2, W_3); 3 levels of nitrogen application (N_0, N_1, N_2) and three of spacing (S_1, S_2, S_3). Since all combinations were represented, this is a factorial experiment; all differential effects can be estimated.

The large-plot treatments (water treatments) were randomized in location in the blocks, and the small-plot treatments in the subplots within the large plots. These small-plot treatments included all 9 combinations of

fertilization and spacing. Field arrangement, however, is not shown in the summary table of data below.

Cochran and Cox (1957) show various short-cuts in analysis of factorials, but only the ordinary methods are shown here. Percentages of forms attacked by bollworms, by combinations of treatments in four blocks, are given in table 30.

The total sum of squares of deviations, $S(x^2) = S(X^2) - [S(X)]^2/n = 111{,}978.05 - (3305.5)^2/108 = 10{,}808.33$.

The large-plot treatment, amount of irrigation water, is summarized by blocks in table 31. Each item is of course the sum of 9 original items; each water total is the sum of 36, each block total the sum of 27 items.

The sum of squares for water treatments: $[(1072.9)^2 + (1031.2)^2 + (1201.4)^2]/36 - (3305.5)^2/108 = 437.22$.

In the same manner the sum of squares for blocks is 1106.53.

TABLE 30. Bollworm damage

Treatments	Block 1	Block 2	Block 3	Block 4	Total
$N_0 S_1 W_1$	16.6	24.4	25.0	15.0	81.0
$N_0 S_2 W_1$	28.8	24.4	43.3	15.5	112.0
$N_0 S_3 W_1$	31.6	25.0	26.6	38.3	121.5
$N_1 S_1 W_1$	25.5	28.8	22.2	22.2	98.7
$N_1 S_2 W_1$	43.3	37.7	35.5	18.8	135.3
$N_1 S_3 W_1$	36.6	40.0	36.6	28.3	141.5
$N_2 S_1 W_1$	26.6	21.6	31.0	27.2	106.4
$N_2 S_2 W_1$	15.5	27.8	30.0	30.0	103.3
$N_2 S_3 W_1$	33.3	65.0	38.3	36.6	173.2
$N_0 S_1 W_2$	22.2	17.8	29.4	20.0	89.4
$N_0 S_2 W_2$	34.4	20.0	31.0	27.8	113.2
$N_0 S_3 W_2$	43.3	41.6	30.0	31.6	146.5
$N_1 S_1 W_2$	19.4	24.4	22.2	12.7	78.7
$N_1 S_2 W_2$	28.8	36.6	30.0	18.8	114.2
$N_1 S_3 W_2$	24.8	45.0	26.6	41.6	138.0
$N_2 S_1 W_2$	28.8	23.3	21.6	25.0	98.7
$N_2 S_2 W_2$	24.4	42.2	32.2	12.2	111.0
$N_2 S_3 W_2$	21.6	35.0	41.6	43.3	141.5
$N_0 S_1 W_3$	18.3	35.5	25.0	31.1	109.9
$N_0 S_2 W_3$	31.0	53.2	27.7	24.4	136.3
$N_0 S_3 W_3$	25.0	40.0	36.6	33.3	134.9
$N_1 S_1 W_3$	13.3	42.8	27.2	40.6	123.9
$N_1 S_2 W_3$	21.0	53.3	20.0	33.3	127.6
$N_1 S_3 W_3$	38.3	43.3	15.0	45.0	141.6
$N_2 S_1 W_3$	18.8	44.4	19.4	32.8	115.4
$N_2 S_2 W_3$	25.5	35.5	35.6	38.8	135.4
$N_2 S_3 W_3$	48.3	43.3	31.6	53.2	176.4
Total	745.0	971.9	791.2	797.4	3305.5

TABLE 31

Water treatment	Block 1	Block 2	Block 3	Block 4	Total
W_1	257.8	294.7	288.5	231.9	1072.9
W_2	247.7	285.9	264.6	233.0	1031.2
W_3	239.5	391.3	238.1	332.5	1201.4
Total	745.0	971.9	791.2	797.4	3305.5

TABLE 32

N — W

	W_1	W_2	W_3	Total
N_0	314.5	349.1	381.1	1044.7
N_1	375.5	330.9	393.1	1099.5
N_2	382.9	351.2	427.2	1161.3
Total	1072.9	1031.2	1201.4	3305.5

S — W

	W_1	W_2	W_3	Total
S_1	286.1	266.8	349.2	902.1
S_2	350.6	338.4	399.3	1088.3
S_3	436.2	426.0	452.9	1315.1
Total	1072.9	1031.2	1201.4	3305.5

N — S

	S_1	S_2	S_3	Total
N_0	280.3	361.5	402.9	1044.7
N_1	301.3	377.1	421.1	1099.5
N_2	320.5	349.7	491.1	1161.3
Total	902.1	1088.3	1315.1	3305.5

The sum of squares for the block-water combinations is $[(257.8)^2 + (294.7)^2 + \cdots + (332.5)^2]/9 - (3305.5)^2/108 = 2767.93$.

The sum of squares for interaction, block × water, is

$$2767.93 - 1106.53 - 437.22 = 1224.18.$$

This is the large-plot error.

A special table is arranged for all N-W combinations, S-W combinations, and N-S combinations (table 32). From each part of the table may be cal-

culated the total sum of squares and the sums for the two main effects; and the interaction sum may be secured by subtraction.

The triple interaction of treatments, N × S × W, is derived by first calculating the sum of squares among all 27 combinations of the treatments. Each item will be the sum of 4 blocks. The columns of totals in the table of original data may be used. From this sum of squares are deducted the six treatment sums included in it: for N, S, W, N × S, N × W, S × W. The remainder is the sum for N × S × W.

The error for the two small-plot treatments (Error B) is a compound of the interactions of these treatments with block × water, and their interactions with blocks. There is no object in separating the components, hence it may be derived by subtracting the sums already calculated from the total. The analysis is summarized in table 33.

No significant differences at all occur except the influence of spacing on bollworm populations. Spacing is highly significant in its influence on bollworm damage. This might be expected; the wider spacings with fewer plant units tend to have a higher rate of bollworm attack, since they are attacked by about the same population. The damage means are: wide spacing, 36.5 percent; medium, 30.2; narrow, 25.1. The experiment had a negative value in showing that modification of prevailing cultural practices could be made with little or no effect. This also applies to yields, not shown here. Some economy was thus shown to be feasible.

TABLE 33

Source of variation	Degrees of freedom	Sum of squares	Mean square
Large plots:			
Between blocks	3	1106.5	368.8
Water treatments	2	437.2	218.6
Block × water treatments (Error A)	6	1224.2	204.0
Small plots:			
Between nitrogen rates	2	189.1	94.5
Between spacings	2	2376.6	1188.3**
Interaction:			
N × S	4	271.2	67.8
N × W	4	161.9	40.5
S × W	4	76.2	19.0
N × S × W	8	284.4	35.6
Error B	72	4681.1	65.0

** Highly significant.

19. Covariance analysis

An example may be found in work on sugarcane borer, in which infestation (percentage count of joints attacked: X) and percentage (measurement: Y) of sucrose were recorded in several experiments at different times and places, with the same variety (table 34). Replications in each experiment totaled six.

Calculation using the whole 36 as a simple series as to correlation and regression, gives $S(x^2) = 6073.59$, $S(y^2) = 88.53$, $S(xy) = +366.20$. The variation in infestation may be divided by analysis of variance, into that between experiments (sum of squares of deviations = 4632.46) and that within experiments (S.S. = 1441.13). The variance for sucrose can be similarly divided into 71.73 between tests and 16.80 within tests. The *products* of deviations can be computed as follows:

Total: $(30.5 \times 14.8) + (53.2 \times 11.8) + \cdots + (23.1 \times 10.5) - (832.2 \times 480.7)/36 = +366.20$.

Between tests:

$$\frac{(239.9 \times 85.7) + \cdots + (69.5 \times 67.4)}{6} - \frac{(832.2 \times 480.7)}{36} = +481.94.$$

Within tests: $(+366.20) - (+481.94) = -115.74$.

The whole may be summarized in table 35.

The "r" computed in each line as $Sxy/\sqrt{S(x^2)S(y^2)}$. It appears that there is a pronounced tendency for the high-sucrose experiments to be more heavily infested, which gives a significant positive correlation between experiments. Within experiments, with sucrose level rather uniform between units, a marked negative correlation shows up. The whole series gives us a significant though moderate positive value, which would be quite deceiving

TABLE 34

Experiment 1		Experiment 2		Experiment 3		Experiment 4		Experiment 5		Experiment 6	
X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
30.5	14.8	39.9	14.6	24.4	15.1	22.9	11.8	20.1	11.4	7.8	12.0
53.2	11.8	22.0	15.3	28.9	14.0	13.7	12.1	7.5	13.1	10.8	10.6
39.2	14.8	40.7	15.2	18.4	14.3	18.8	12.5	7.2	13.3	1.8	12.0
36.8	14.6	33.7	15.5	15.4	15.6	15.8	12.6	13.4	11.9	10.5	11.4
44.0	14.8	40.3	14.7	15.3	15.5	13.1	12.4	6.8	12.5	15.5	10.9
36.2	14.9	40.2	15.2	37.7	13.7	13.6	12.8	13.0	12.5	23.1	10.5
239.9	85.7	216.8	90.5	140.1	88.2	97.9	74.2	68.0	74.7	69.5	67.4

Grand totals: X, 832.2; Y, 480.7.

TABLE 35

Source of variation	Degrees of freedom	$S(x^2)$	$S(xy)$	$S(y^2)$	r	d.f. for r
Total	35	6073.59	+366.20	88.53	+0.50	34
Between experiments	5	4632.66	+481.94	71.73	+0.84	4
Within experiments (error)	30	1441.13	−115.74	16.80	−0.74	29

TABLE 36

Source of variation	Sum of squares of errors of estimate (S.S.E.)	Degrees of freedom	Mean square
Total	66.45	34	—
Error	7.50	29	0.26
Difference	58.95	5	11.79

by itself. The within-experiment or error calculation is the better guide to the underlying relation, free of experiment differences. If the difference in sucrose between experiments was being critically investigated, we might wish to correct for the underlying correlation. This is most accurately done by making the "F" test with residual variance, as shown below. Errors of estimate are calculated by the short-cut forumla, $S(y^2) - [S(xy)]^2/S(x^2)$, or $(1 - r^2) \cdot S(y^2)$. Results are shown in table 36. The mean square 11.79 is tested against 0.26 as error. Thus differences between experiments are shown to be more marked than would appear without allowing for the within-experiment correlation. The test may be supplemented by comparing *adjusted* class means of sucrose content, using the formula $Y - bx$. The "b" used is the regression coefficient, $S(xy)/S(x^2)$, and is calculated from the *error* classifications; "x" in this case will be the deviation of the class mean from the general mean. E.g., general mean $\bar{x}$ is 832.2/36 or 23.1; in Experiment 1, $\bar{x} = 40.0$; deviation, $\bar{x} - \bar{\bar{x}} = 40.0 - 23.1 = +16.9$; $b = -115.74/1441.13 = -0.0803$; $bx = -0.0803 \times 16.9 = -1.4$; $Y - bx = 14.3 - (-1.4)$ or 15.7.

Covariance may be analyzed in even more complex experiments by these methods. In the randomized block analysis, if two variables had been measured in each unit, we could separate covariance for varieties, blocks, and error. In special adjustments such as the one just above, we use only those parts of the analysis which are of special interest. We add "treatment" and "error" sums of squares and products, to form a new "total." "Block" influences have been evaluated and set aside.

20. Covariance analysis, randomized block experiment

In a nut insect problem, fear was expressed that the varying size of trees over the experiment would interfere with recognition of treatment differences. The cross-section area of trees was taken for each of the single-tree plots. There were 5 treatments replicated 10 times in a randomized block arrangement.

In table 37, blocks are indicated by letters A to J, plot yields (Y) are the upper figures in each cell, cross-section area (X) are the lower figures. Correction factors (C.F.) are as follows:

Cross-section area (X), $(4535.3)^2/50 = 411{,}378.92$
Yield (Y), $(1419.7)^2/50 = 40{,}310.96$
Covariance (XY), $(4535.3 \times 1419.7)/50$; 128,775.31

Totals for $S(x^2) = 19{,}237.87$; $S(xy) = 3{,}889.25$; $S(y^2) = 8070.69$ are calculated as in example 19 or as in a simple correlation problem.

$S(x^2)$ for blocks is $[(356.0^2 + (415.7)^2 + \cdots + (475.6)^2]/5 - \text{C.F.} = 5119.20$.

$S(x^2)$ for treatments is similarly computed as 2563.85.

$S(x^2)$ for error (interaction, block × treatment) is $19{,}237.87 - 5119.20 - 2563.85 = 11{,}554.82$.

$S(y^2)$ for blocks, treatments, and error is figured in the same way as 4528.89; 558.17; 2983.63, respectively.

TABLE 37

Treatment	A	B	C	D	E	F	G	H	I	J	Totals
1:											
Y	27.1	48.5	55.0	38.4	19.8	25.5	28.1	21.4	40.6	14.4	318.8
X	74.5	113.0	83.5	113.0	66.9	71.6	66.0	96.2	88.8	133.8	907.3
2:											
Y	13.0	28.3	52.5	38.3	15.3	47.0	23.3	22.9	13.4	10.9	264.9
X	41.7	69.7	99.7	88.8	85.6	114.9	92.5	86.7	85.6	69.3	834.5
3:											
Y	30.5	28.3	54.5	23.7	15.9	10.7	21.0	32.7	42.4	11.5	271.2
X	73.5	93.0	111.3	71.6	87.2	119.6	121.5	94.1	129.9	98.6	1000.3
4:											
Y	40.6	31.9	65.8	37.4	15.5	25.9	31.7	40.2	18.3	18.9	326.2
X	88.8	72.1	114.9	88.8	69.3	118.4	100.3	112.4	103.1	104.2	972.3
5:											
Y	29.7	27.0	29.9	30.2	19.0	26.5	17.6	22.1	16.0	20.6	238.6
X	77.5	67.9	74.5	92.0	86.1	110.7	58.0	92.0	92.5	69.7	820.9
Total:											
Y	140.9	164.0	257.7	168.0	85.5	135.6	121.7	139.3	130.7	76.3	1419.7
X	356.0	415.7	483.9	454.2	395.1	535.2	438.3	481.4	499.9	475.6	4535.3

$S(xy)$ for blocks is $[(140.9 \times 356.0) + \cdots + (76.3 \times 475.6)]/5 - \text{C.F.}$ or +768.96. For treatments, $S(xy)$ is $[(318.8 \times 907.3) + \cdots + (238.6 \times 820.9)]/10 - \text{C.F.}$ or +686.55. For error, $S(xy)$ is (+3889.25) − (+768.96) − (+686.55), or 2433.74.

The summary is shown in table 38.

The influence of cross-section area is seen to be small, though correlation if computed from the "error" line of the table will be found to be significant. The treatment differences in yield may be tested without allowance for correlation. The mean squares are: treatment 558.17/4 or 140; error 2983.63/36 or 83. F is 140/83. By the adjustment procedure the error mean square is reduced from 83 to 71, but the treatment mean square is reduced even more. The treatments do not show significant differences in either case. The indication is that part of the apparent tendency to difference in treatments was due to the correlation and to unequal distribution of tree size among treatments.

Covariance analysis will reduce the error and permit closer evaluation of differences, if correlation is real. It may enable detection of differences which would escape ordinary analysis, or may show that apparent differences are not real after allowance for correlation.

21. A problem involving the probit transformation

Use of transformations in regression, and of probits, is touched upon in Chapter 1. Use of probits assumes a "normal" distribution of logs of susceptibility, and this seems to be close to the truth in many cases. If a straight line is fitted to logs of concentration and probits of mortality, it can be used to estimate the "LD-50", or concentration effective on 50 %, one of the best measures of comparison; also to predict mortality at a given concentration. It also estimates the distribution of logs of susceptibility among subjects (that is, logs of the amount just sufficient to kill each animal). The log LD-50 estimates the mean; the regression coefficient, the reciprocal of the standard deviation. Bliss (1935a and 1935b) and Finney (1952) give details of procedure.

TABLE 38

Source	D.F.	$S(x^2)$	$S(xy)$	$S(y^2)$	D.F., adjusted	S.S.E.E.	M.S.
Total	49	19,237.87	+3889.25	8070.69	—	—	—
Blocks	9	5,119.20	+768.96	4528.89	—	—	—
Treatments	4	2,563.85	+686.55	558.17	—	—	—
Error	36	11,554.82	+2433.74	2983.63	35	2471.02	70.60
T & E	40	14,118.67	+3120.29	3541.80	39	2852.20	—
Difference	—	—	—	—	4	381.18	95.30

Data on house fly mortality from pyrethrum spray may be used to illustrate calculations.

Concentration, milligrams per liter (mg./l.)	250	500	1000	2000
Response	11/140	40/142	110/134	132/136

Several degrees of complexity are possible. First, a simple graphic solution can be made. Logs of concentration are plotted on the horizontal axis, probits of mortality on the vertical, as in Chapter 1. A line is fitted by eye and the log concentration for 50 % (5 probits) is read off. This quick method often gives surprisingly good results, and is used in starting out the more complex solutions. Probits for different percentages are given by Finney (1952), Fisher and Yates (1963), and others. Probits can also be secured from any table of areas of the normal frequency curve, such as found in most statistical textbooks. Percentage is taken as the proportion of the area reading from the middle (50 %); the probit is the corresponding plus or minus standard deviation value, with *5* added.

For example, the second response above is 40/142, which is 28.2 % or 0.282. This is 0.218 below 0.50. In Snedecor's cumulative normal table (1956, Ch. 8), 0.218 is reached at 0.58 standard deviation ("t") from the mean. This is below the mean, so it is −0.58; adding 5, we have 4.42 for the probit. The third response is 110/134 or 0.821; it is 0.321 above 50 %, with standard deviation value of +0.92 and probit 5.92.

Second, a simple regression line can be fitted as in Chapter 1, instead of using an eye-fit. If mathematical fitting is used, however, it is better to use weights, as the probit values near 0 and 100 % are not so well determined and need a lower weight.

Third, these weights can be used, and a weighted linear regression fitted as in the example below, but without the "maximum likelihood" adjustments of the probits. Weights for various probits can be secured from tables in sources named above.

The weight (reciprocal of theoretical variance) is nz^2/pq, where z is the ordinate of the normal curve, p is the proportion, and q is $1 - p$. The part z^2/pq is called the weighting coefficient ("w"); it can be secured from a table or by calculation, and multiplied by n. The first probit calculated, 4.42, is found in Snedecor's table of normal ordinates (1956) to have an ordinate z of 0.337, corresponding to the 0.58 standard deviation or "t". The p is 0.282, the q is 0.718, the pq 0.2025, and "w" $= z^2/pq$ is 0.56. For $n = 142$, the weight is 0.56×142 or 79.52. Weights are usually secured for probits off the provisional line, rather than the actual percentages.

The fourth stage is the full-scale probit solution, using adjusted probit values as well as weights. The adjustments are aimed to provide a slightly

better estimate of mortality, allowing for the trend among all observations. They were first used for cases of 100 % mortality in small samples, where we are sure 100 % does not represent population conditions. They have since been applied to all probits, but in good data the change from original probits is not great.

These adjusted or working probits are calculated as

$$y = Y + \left(\frac{p - P}{z}\right)$$

where y is the working probit, Y is the expected probit read off a provisional line, p is the original proportion observed, P is the proportion corresponding to Y (from table) and z is the ordinate corresponding to Y. The calculation is laborious, and tables to eliminate calculation are bulky and are not very widely available outside of Finney's books. He shows very useful tables. One of these books is almost indispensable for such calculation. Where electronic machines are used we make them calculate their own weights, adjustments, etc.

In the full-scale analysis we first get logs and probits, plot and fit a line by eye. From this line we read "expected" probits, and for these expected probits we secure weights and working probits. The calculation is then carried out as a weighted linear regression. In some cases if the new equation gives probits much different from the expected probits these new probits are taken as new expected values, and the calculation repeated till stability is reached. Many research organizations with access to an electronic calculator have a "probit program," which is a boon for these heavy calculations.

The calculation will be carried through for the example shown. The usual set-up of many columns and few rows is switched for convenience in table 39. We calculate $\bar{x}$ as 606.828/217.44 or 2.79078; $\bar{y}$ likewise as 4.95957. It is best to carry them to 5 places at this stage. The weighted sums of squares and products of deviations are calculated: $S[nw(X - \bar{x})^2] = S(XnwX) - \bar{x}S(nwX)$; $S[nw(y - \bar{y})^2]$ likewise. $S[nw(X - \bar{x})(y - \bar{y})] = S(Xnwy) - \bar{x}S(nwy)$. The order of multiplication might be reversed in the case of the products. These sums of squares and products can be run up on a desk calculator without entering individual products.

Calculating thus, we have:

$$S[nw(X - \bar{x})^2] = 1708.8876 - (2.79078 \times 606.828) = 15.3642$$

$$S[nw(X - \bar{x})(y - \bar{y})] = 3069.7146 - 3009.5995 = 60.1151$$

$$S[nw(y - \bar{y})^2] = 5588.3140 - 5348.4400 = 239.8740$$

TABLE 39

					Total
Concentration (mg./liter)	250	500	1000	2000	—
X (log. concentration)	2.4	2.7	3.0	3.3	—
n	140	142	134	136	—
r (number killed)	11	40	110	132	—
p	.079	.282	.821	.971	—
Empirical probit	3.59	4.42	5.92	6.90	—
Y (expected probit, from graph)	3.6	4.7	5.8	6.9	—
w	.30	.62	.50	.15	—
nw	42.00	88.04	67.00	20.40	217.44
y (working probit)	3.59	4.44	5.91	6.90	—
nwy	150.780	390.898	395.970	140.760	1078.408
nwX	100.800	237.708	201.000	67.320	606.828

Part accounted for by regression = $(60.1151)^2/15.3642 = 235.2108$

Remainder, chi-square in this special case 4.6632

This is not significant, with 2 degrees of freedom (number of concentrations minus two).

This shows that variation can be ascribed to ordinary random chance, and that the theoretical variance is adequate. Regression coefficient, b, = $60.1151/15.3642 = 3.9127$. $a = \bar{y} - b\bar{x} = -5.96$. Equation is $Y = -5.96 + 3.9127\ X$. Using this equation we calculate new expected Y's from the $4X$'s; they are 3.43, 4.60, 5.78, 6.95. These do not vary much from the expected Y's used above, and another cycle is probably not needed. The new Y's could be used to recalculate the last 6 lines of the table if it seemed needful.

Setting $Y = 5$ in the equation, X solves as 2.80; this is the log LD-50 or m, and the LD-50 is estimated as about 630. The variance of m is $1/b^2[1/Snw + (m - \bar{x})^2/Snw\ (X - \bar{x})^2]$. This is a modification of a general regression formula, utilizing the fact that theoretical variance is embodied in the weights. We calculate $V_m = 0.000304$ and $S_m =$ about 0.017. The variance of b is estimated as $1/[Snw(X - \bar{x})^2]$; like V_m, this is a special modification of a general formula. In our problem V_b is 0.0651 and s_b about 0.255.

Where chi-square is not significant, we assume *infinite* degrees of freedom (since these are theoretical variances) and use a t-multiplier of about 2. If chi-square *is* significant, we should multiply the variances by (chi-square divided by its degrees of freedom); and in using the standard errors, assume only *2* d.f. (t multiplier 4.3).

There are further refinements in analysis, procedure for comparing curves, other short-cuts, etc., shown by Finney (1952); but the above account presents most of the special features of probit analysis.

22. A problem using the probit-log transformation in an approximate comparison of insecticides

Three insecticides were to be compared in a laboratory study of toxicity by test insects in a standard manner. The best comparison of insecticides is that of the strengths required for a given effect. The concentration giving 50% mortality gives the closest comparison. This strength is best estimated from a dosage mortality curve by interpolation. The insecticides were each tested at three concentrations on six days, against an adequate sample of insects. The three concentrations gave opportunity to determine an approximate dosage-mortality curve each time. For instance, test in insecticide C on the first day showed results as follows:

Concentration units	Mortality (%)
1	25
2	42
4	73

This could be done by the mathematical methods already shown, but a graphic method was found to give sufficiently accurate results.

The probit-log transformation gives approximate linearity, making interpolation easy. A special plotting paper, with logarithmic intervals on the horizontal axis and probability intervals on the vertical axis, is available. On this, the original readings can be plotted, and the probit-log relation will appear. A straight line can be fitted by eye, and the concentration for 50% mortality read off. When 50% concentrations were read off they were tabulated (table 40). The days, involving different groups of insects differing in susceptibility, and other uncontrolled variation, correspond to blocks in a randomized block field test. The same analysis is adapted. It was carried out and summarized in table 41. This shows that the insecticides really differed. A computation of "least significant difference" indicates that A and B are not significant in their difference, but that C is significantly lower than either. As lower concentrations are required for 50% kill, it is indicated as more potent.

TABLE 40

Insecticide	Day						Total
	1	2	3	4	5	6	
A	3.6	2.5	1.3	2.2	2.2	1.6	13.4
B	3.2	1.8	1.3	2.2	2.1	1.3	11.9
C	2.2	1.6	1.5	1.3	1.8	1.2	9.6
Total	9.0	5.9	4.1	5.7	6.1	4.1	34.9

TABLE 41

Source of variation	Degrees of freedom	Sum of squares	Mean square
Days	5	5.37	1.07**
Compounds	2	1.22	0.61*
Interaction, days × compounds (error)	10	1.01	0.10

** Highly significant. * Significant.

The *logarithms* of LD-50's may be used instead of concentrations in analysis; they will be somewhat better adapted to the assumptions of analysis, and may give more precise results than simple concentrations. However, in the present case, their use leads to the same conclusions.

The solution may of course be made by more complex methods, such as those used by Finney (1952). However, such an approximate solution as that above gets most of the information. In this particular case more complex methods were carried out with no change in decision.

23. Determination of synergism by short-cut methods

Synergism in a mixture of insecticides is defined as some interaction giving a greater effect than could be expected from their separate actions. It has been mentioned in a preceding chapter. The writer has published a discussion of synergism (Wadley 1945). The maximum effect of a mixture, short of synergism, is usually that of similar, additive effect (Bliss 1939). To determine synergism it is necessary (1) to define this additive effect from separate action, and (2) to show that the mixture gives results greater than the additive effect.

Finney (1952) shows detailed analytical procedure based on probit analysis, for this problem. A shortened graphic procedure is usable in this connection. Trials must be available giving a dosage-mortality curve for each ingredient alone, in the middle and upper range of mortality. The amount of each necessary to produce a given mortality will give an estimate of *equivalence*. Using this estimate, the concentration of the mixture can be expressed in terms of either one of the ingredients. The expected mortality can be determined from the dosage-mortality curve for the ingredient chosen. This expected mortality can be compared with actual mortality for the mixture.

For plotting, concentration and mortality can be transformed to logs and probits, respectively, giving a more linear curve. However, log-probability paper is available, on which dosage and percentage can be plotted without transformation for graphic work, with the same effect.

Data are taken from an article by Martin (1942), also used by Finney (1952). They are mortalities of an aphid species treated with rotenone and deguelin (table 42). These data were plotted on the log-probability paper mentioned, and straight lines were fitted by eye. The 50% mortality is estimated from the line to require a 4.8 concentration of rotenone or 13.2 of deguelin. Thus deguelin appears to be 0.36 as good as rotenone. At 90%, the comparison is 9.7 to 28.0, or a ratio of 0.35. The average ratio is 0.355.

Mortalities with mixtures of the two are shown in table 43. After calculation of rotenone equivalent, the expected mortality is read off the rotenone line already drawn. The actual exceeds the expected mortality slightly at each point, but not greatly. They agree fairly well. Martin's data were based on replicated trials, and he cites standard errors that show the excess mortality is not significant.

Where replicated trials are made, repeated determinations of equivalence can be made and standard errors calculated for both expected and actual effects; significance would thus be tested.

Suppose that the mortalities above were drawn from trials replicated on 3 days (table 44).

We could determine expected equivalence for each day, by the process outlined above. When this is done, rotenone equivalents for deguelin are 0.36, 0.36, 0.35 respectively; little variation appears.

Suppose further that the mixtures yielded results on the *same* days as in table 45. The same rotenone equivalent is used for all 3 days, since the values were so similar each day. Had they differed much, the figure for each

TABLE 42

Rotenone		Deguelin	
Concentration (mg./liter)	Mortality (%)	Concentration (mg./liter)	Mortality (%)
3.8	33.3	10.1	37.5
5.1	52.2	20.2	70.8
7.7	85.7	30.3	95.9
10.2	88.0	40.4	94.0

TABLE 43

Rotenone (mg./liter)	Deguelin (mg./liter)	Rotenone equivalent (rotenone + 0.355 deguelin)	Actual mortality (%)	Expected mortality (%)
2.0	8.1	4.9	58.7	50
3.0	12.2	7.3	79.2	78
4.0	16.3	9.8	93.5	90

TABLE 44

Rotenone				Deguelin			
Concentration	Mortality			Concentration	Mortality		
	Day A	Day B	Day C		Day A	Day B	Day C
3.8	33	29	38	10.1	35	37	40
5.1	48	50	58	20.2	70	67	75
7.7	85	83	89	30.3	96	95	97
10.2	86	89	90	40.4	89	94	98

TABLE 45

Concentration			Mortality (expected in parentheses)		
Rotenone	Deguelin	Rotenone equivalent	Day A	Day B	Day C
2.0	8.1	4.9	55(49)	57(48)	64(56)
3.0	12.2	7.3	79(76)	76(76)	83(83)
4.0	16.3	9.8	92(89)	93(90)	95(93)

day might have been used for that day's results. The dosage-mortality curve for rotenone for each day is taken; the mortality corresponding to rotenone-equivalent concentration is read off (for 4.9, 7.3, 9.8) as *expected.* Thus for each concentration level we have an expected and actual series for comparison by standard error methods. Using pair differences, significance seems to be attained in the higher and lower concentrations of this hypothetical series, but not in the middle concentration. If mixtures were not tested on the same days as the separate ingredients, pair differences could not be used, and error would likely be larger.

It should be stated that synergism, in the strict sense, is limited to mixtures of substances *each* having some toxicity. Where a nontoxic substance improves that action of the toxicant, it is not synergism but simple activation, and statistical treatment is simpler. Significant increase in mortality in replicated trials should be sufficient to show simple activation.

24. Analysis of experimental error

Where the error estimate is *compound*, the contribution of major and minor sources can be evaluated. In an experiment where the major units are plots, and plot values are derived from sampling; or a sampling investigation with fields as units, and subunits within fields; this will apply. It is necessary for this evaluation that randomness be observed in arrangement of both units and subunits. (For ordinary analysis based strictly on units, randomness is not needed among subunits.)

An abridged example has already been given in the chapter on sampling, Section 7. As another example, spray deposits were measured at 3 times with with several determinations per time. The 3 times themselves may be thought of as a sample from a large series of times that might be taken. The determinations each time are then in the category of subsampling. The determinations within times, and the times themselves, are essentially random as far as we know. The analysis is summarized:

	Degrees of freedom	Mean square
Between times	2	21.38
Between determinations within times	12	1.69

Using the notation of Chapter 2, Section 7, $n = 3$, $k = 5$, $B = 1.69$, $A =$ (mean square between times minus mean square within times), divided by k. Thus $A = (21.38 - 1.69)/5$, or $19.69/5$, or 3.94. A is the variance between times, over and above that to be expected from within times, brought to a single-determination level. If there is *no* real variation between times at all, we would expect some variance (about 1.69 more or less) between times, because within-time variance would produce it. The significant mean square for between times shows that there *is* some real variance, over and above the expected 1.69. Thus two sources of variation affecting any single determination are separated.

For such a single determination, at any given time, we may expect a variance around the true mean of $A + B$, or $3.94 + 1.69$, or 5.63. If two determinations are made at one time, the expected variance of their mean is $3.94 + 1.69/2$, or 4.78. The "A" part will not be reduced, unless time is replicated. If 4 determinations are made at a single time, the variance estimate is $3.94 + 1.69/4$ or 4.36. Suppose, however, 2 determinations are made at each of 2 times, 4 in all; the estimate is $3.94/2 + 1.69/4$ or 2.39.

By this process, expected variances (and standard errors) of means can be worked out, and more efficient combinations can often be derived in compound sampling. If A has a real *existence* (indicated by significant "F"), a gain will always be given by splitting up the sampling. The limit to the process will be practical, not statistical. Experimental costs should be considered (see Chapter 2, Section 7).

This process may be used in exploring effects of number of samples and subsamples; or of number of replications and of sample units per plot. It may be used to test effects of size of plots, where a plot is thought of as composed of several adjacent units.

25. Gain from spreading sampling

This has been briefly treated in Chapter 2, Section 7, in a case where 20 fields and 5 units per field were taken to estimate wheat yield in a county.

Another case is taken from insect population sampling. In cotton fleahopper counts, 100 terminals were taken at each of 2 points per field, in a number of fields. The number of insects per terminal was counted. It was desired to investigate possible improvement from studying more points per field and fewer terminals per point. For this purpose each set of 100 may be regarded as 4 adjacent sets of 25. While lots of 25 terminals were not counted separately, and no determinations of adjacent units were available, the theoretical minimum variance was used. It seemed likely that adjacent samples would show a variance not much above this theoretical value. These population counts would be expected to have the Poisson variance equal to the mean.

The average variance among counts between locations within fields was 44.3. The variance of the mean of 2 points was 22.2. The theoretical variance was 17.4. We may put these on a basis of per 25 instead of per 100 terminals. If the variance of numbers per 100 is 44.3, the variance of numbers per 25 is estimated as 11.08; but the variance of numbers per 100 estimated from only 25 is 16 times as great, or 177.2. In the same way, the theoretical variance per 100 is 17.4; per 25, 4.35; per 100 estimated from 25, 69.6. This can be set up: $kA + B = 177.2$, $k = 4$ (since 4 sets of 25 occur in 100), $B = 69.6$, $A = 26.9$. The variance of the mean of n locations and k sets of 25 per location is $A/n + B/nk$. For 1 set of 25, at 1 point, it is $26.9/1 + 69.6/1 = 96.5$ for estimated population per 100 terminals.

For 1 point with 4 sets of 25, it is $26.9/1 + 69.6/4 = 44.3$

For 2 points with 4 sets of 25 each, it is $26.9/2 + 69.6/8 = 22.1$

For 4 points with 1 set of 25 each, it is $26.9/4 + 69.6/4 = 23.4$

For 8 points with 1 set of 25, it is $26.9/8 + 69.6/8 = 12.1$.

The 2×4 is the combination that had been used; 4×1 is nearly as good with half the material; while 8×1 is considerably better with the same amount of material examined.

This scheme may be extended to more than two orders of compound sampling. In a grasshopper egg-sampling technique study, examinations were made of 10 fields, 50 locations in each field, and 2 adjacent subunits at each location. Analysis of variance among egg-pod counts is as shown in table 46. It will be seen that under these particular circumstances, adjacent subunits varied nearly as much as separate locations. The variance of the latter reached significance only because of the large numbers of observations. The theoretical variance was 0.19. The variance of a single observation may here be subdivided into 3 portions, which may be called A, B, and C. The mean square between subunits within locations, 0.20, is C. If the number of fields is n, the number of locations per field is k, and the number of subunits per location is j; the mean square between locations within fields is $jB + C$. This gives us $2B + C = 0.26$; $2B = 0.06$; $B = 0.03$. Also, the

TABLE 46

	Degrees of freedom	Mean square
Between fields within the area	9	1.16**
Between locations within fields	490	0.26*
Between subunits within locations	500	0.20

** Highly significant. * Significant.

TABLE 47

	Degrees of freedom	Mean square
Between blocks	3	1378.98**
Between treatments	9	397.55**
Interaction, block × treatment (error for treatment)	27	81.37**
Between trees within plots	80	19.73

** Highly significant.

mean square between fields is $kjA + jB + C$; or $100\,A + 2\,B + C = 1.16$. Solving, $A = 0.009$, or about 0.01. The variance of the mean of any combination of n, k and j will be $A/n + B/nk + C/nkj$.

26. Analysis of error in a plot experiment

An experiment testing sprays on apples against the codling moth had 10 treatments, 4 replications in randomized blocks, and 3 trees per plot. Data were estimates of worms per 100 apples. Analysis was summarized as in table 47. The mean square between trees within plots is in a sense error for interaction; it is also a component of this interaction, which is error for treatments. Since the 3-tree plots are randomly chosen for treatments within each block, the within-plot differences are an expression of random differences among adjacent trees. The three trees may be looked upon as 3 adjacent plots. It should be obvious that where treatments must be in somewhat separated plots, variance among adjacent plots is not adequate as error. It may be used, however, to test effect of size of plot within reasonable limits.

In this case, mean square between trees within plots is an estimate of B. The mean square for interaction, block × treatment, is an estimate of variance among plots treated alike within a block. This is $kA + B$. Since $k = 3$ (trees per plot), $3A + 19.73 = 81.37$, and $A = 61.64/3$ or 20.55. The error mean square for n replications and k trees per plot will be estimated as $A/n + B/nk$, where the plot size is reasonably small. In this case, solution led to a change to single-tree plots; since they made better use of

the total material than 3-tree plots, and were satisfactory from the practical standpoint. It will be seen that single-tree plots replicated 12 times give a lower error variance than 3-tree plots with 4 replications. As a matter of fact, 6 replications of single-tree plots give in this variable material about as good results as 4 replications of 3-tree plots. The variance of a treatment mean in the experiment as carried out is 81.37/12, or 6.78, since the trees are units in analysis and there were 12 per treatment. This is the same as 20.55/4 + 19.73/12. For 6 replications, single-tree plots, the estimate is 20.55/6 + 19.73/6, or 3.42 + 3.29, or 6.71, with half the trees.

As stated, this process may also be applied to cases where plot values are determined by sampling. It will give a test of effect of varying combinations of sample units and plot replication. To use the process validly, samples should be random within the plots. Randomness within plots is not required for practical evaluation of experimental results, if the plots themselves are assigned randomly.

27. Solution for minimum error of compound sampling with fixed costs

This subject has been touched upon in Chapter 2, Section 7. An example may be drawn from grasshopper egg survey work. In the preliminary attack on the problem, cost estimates were limited to time used, which could be converted to money cost fairly easily. The "overhead cost" of locating a field, writing it up, etc., was estimated as 60 minutes. The cost per unit after reaching a field was estimated as 4 minutes. The total cost allowable was taken as 720 minutes. For one county the mean square between fields (using egg-pods per unit area) was 0.93; between units within fields, the mean square was 0.42. The variances A and B were estimated as before. Since k, the number of units per field, was 10; $10A + B = 0.93$; $B = 0.42$; $A = 0.05$. If C = cost per unit (4 minutes), CD = cost per field (60 minutes); optimum $k = \sqrt{BCD/AC}$. Here, estimated

$$k = \sqrt{(0.42 \times 60)/(0.05 \times 4)} = \sqrt{25.20/0.20} = \sqrt{126.00} = 11.2.$$

The optimum for k under these conditions is about 11. Then $n = T/(CD + kC)$, where T is the limiting cost. This gives us $720/[60 + (11 \times 4)]$, or 720/104, or about 7. The estimate of optimum combination for low error with 720 minutes work is 7 fields with 11 units per field, in this set of conditions.

It is true that overhead cost per field would be lower if number of fields was greatly increased, but the above approximation should work over a reasonable range. It can be seen that this same relation could be used in another way. The variance of the mean could be set at a satisfactory figure, and solution made for the n and k needed to give minimum total cost with

this variance. In such a solution, k will be estimated as before, and n as $(kA + B)/(V_{\bar{x}}k)$, where $V_{\bar{x}}$ is the desired variance of the mean.

28. Factorial design and single degrees of freedom

This subject has been mentioned in Chapter 3; the advantage in efficiency, and some specific designs, have been discussed. The cage experiment with 4 criteria of classification (Example 17) is a $2 \times 3 \times 3$ factorial, as far as experimental treatments are concerned. The 4 years are treated as replications or "blocks." The split plot experiment (Example 18) is a $3 \times 3 \times 3$ factorial, with the complication that one treatment is on a whole-plot scale, the others in subplots.

An experiment a few years ago on cotton insects included treatments with cryolite and sulphur, with all combinations of presence or absence. There were three replications in randomized blocks. The percentages of bolls injured by bollworm are given in table 48. The treatment totals are as shown in table 49. This may readily be analyzed by the ordinary method, as summarized in table 50.

The treatment degrees may readily be subdivided by ordinary methods into two effects, and interaction of the two. The sum of squares for the treatments is 287.39. Of this, the sum for the degree of freedom comparing sulphur and no sulphur is $[(92.3)^2 + (85.5)^2]/6 - 2634.40$ (the correction factor). This comes out as 3.86, with one degree of freedom. The sum for cryolite vs. no cryolite is $[(59.9)^2 + (117.9)^2]/6 - 2634.40$, or 280.34. The

TABLE 48

	Block 1	Block 2	Block 3	Total
No treatment	19.7	23.5	19.0	62.2
Sulphur alone	17.7	22.5	15.5	55.7
Cryolite alone	8.2	11.2	10.7	30.1
Cryolite and sulphur	11.2	8.8	9.8	29.8
Total	56.8	66.0	55.0	177.8

TABLE 49

	Sulphur	No sulphur	Total
Cryolite	29.8	30.1	59.9
No cryolite	55.7	62.2	117.9
Total	85.5	92.3	177.8

TABLE 50

	Degrees of freedom	Sum of squares	Mean square
Blocks	2	17.41	8.70
Treatments	3	287.39	95.80
Error	6	28.02	4.67

TABLE 51

Degree of freedom	Reading for—												Net sum	Divisor	(Sum)²/ divisor
	Block A				Block B				Block C						
	o 19.7	s 17.7	c 8.2	sc 11.2	o 23.5	s 22.5	c 11.2	sc 8.8	o 19.0	s 15.5	c 10.7	sc 9.8			
B_1 (blocks)	+	+	+	+	−	−	−	−	0	0	0	0	−9.2	8	10.58
B_2 "	+	+	+	+	+	+	+	+	−2	−2	−2	−2	+12.8	24	6.83
T_1 (Sulphur)	−	+	−	+	−	+	−	+	−	+	−	+	−6.8	12	3.85
T_2 (Cryolite)	−	−	+	+	−	−	+	+	−	−	+	+	−58.0	12	280.33
T_3 (S × C)	+	−	−	+	+	−	−	+	+	−	−	+	+6.2	12	3.20
$B_1 \times T_1$	−	+	−	+	+	−	+	−	0	0	0	0	+4.4	8	2.42
$B_1 \times T_2$	−	−	+	+	+	+	−	−	0	0	0	0	+8.0	8	8.00
$B_1 \times T_3$	+	−	−	+	−	+	+	−	0	0	0	0	+6.4	8	5.12
$B_2 \times T_1$	−	+	−	+	−	+	−	+	+2	−2	+2	−2	+6.4	24	1.71
$B_2 \times T_2$	−	−	+	+	−	−	+	+	+2	+2	−2	−2	−16.0	24	10.67
$B_2 \times T_3$	+	−	−	+	+	−	−	+	−2	+2	+2	−2	−1.6	24	0.11

sum for their interaction is 287.39 − 280.34 − 3.86, or 3.19. Practically all the treatment variance is due to effect of cryolite.

All "degrees of freedom" in this simple set-up may easily be separated. With more than two degrees in a set, more than one set may be arranged from the same material. The 3 treatment degrees are arranged to give sulphur vs. no sulphur, cryolite vs. no cryolite, and interaction of cryolite and sulphur. While other arrangements might be made, this is the one that brings out the interesting comparisons. The two degrees for blocks offer no especially interesting comparisons; they may follow the simplest scheme, A − B and A + B − 2C. The 6 degrees for interaction, block and treatment, may be derived as products of the corresponding treatment and block degrees. In table 51 the coefficients are negative or positive; where a coefficient is 1 the figure is omitted, using a simple + or −. The divisors are the sum of squares of coefficients. Treatment combinations are indicated by o, s, c, sc

Selection of combinations follows several simple rules. The coefficients for any single degree must total zero; the products of corresponding coefficients in any two degrees must add to zero. Any single value should not be

used by itself more than once; it may be used later in combinations, but not more than once in the same combination (note the block comparisons, (A − B), (A + B − 2C)). There are only $n - 1$ independent comparisons in a set. "Net sum" is found by adding horizontally the individual values at the column heads, according to coefficients. For B_1, for example, 19.7 + 17.7 + 8.2 + 11.2 − 23.5 − 22.5 − 11.2 − 8.8 = −9.2 (table 51). The sum of squares for the d.f. is (net sum)²/divisor. The divisor is the sum of squares of coefficients.

The sum of squares for each individual degree is shown in the last column. It will be observed that the sums for the 2 "block" degrees, and the sums for the 6 "error" or interaction degrees, add up to the totals already derived for blocks and error. The sums for the treatment degrees also check with the previous calculation, with small differences due to rounding.

The block and error degrees of freedom, in an experiment like this, ordinarily need not be separated. The individual treatment degrees are often of considerable interest. They may be separated by a short-cut version of the process above. The little 2 × 2 table of cryolite vs. no cryolite, sulphur vs. no sulphur, may be utilized. The difference of column totals, 85.5 and 92.3, will give the net sum −6.8 for sulphur as in the detailed table. The difference of row totals, 59.9 and 117.9, will give −58.0 for cryolite. The difference of *diagonal* totals, 92.0 and 85.8, will give the net sum for interaction of sulphur and cryolite. Using the proper divisors the sums of squares may easily be calculated.

29. Confounding in a plant disease problem

An experiment in bulb growing was run as a 3 × 3 × 3 factorial. One treatment, A, was prestorage fungicidal treatment; at 3 levels, none, weak, and strong. The two latter were nearer equal than would be supposed, since the stock solution used for the "weak" was more potent. The second treatment, B, was drying; none, 2 hours, 12 hours. The third, C, was preplanting fungicidal treatment; none, weak, or strong.

The 27 combinations were made up in 3 groups of 9 each, so as to confound 2 of the 8 degrees of freedom for triple interaction with groups. The groups were then randomized in 3 blocks of 9 small plots each. There was no absolute replication. The triple interaction was relied upon as an estimate of error. The criterion was the number of diseased bulbs produced by the different treatments in the standard plantings. The numbers are rather small for analysis without transformation, and the experiment is limited in scope, but it will serve to illustrate the methods.

The arrangement of groups had to be made with great care. As with most of these designs, this arrangement is a paper problem, quite distinct from

the later problem of randomizing the actual field locations. Each of the 27 units was classified according to the level of three treatments. These may be called A_1, A_2, A_3; B_1, B_2, B_3; C_1, C_2, C_3. The treatment involving no prestorage or preplanting fungicide, and no drying, is A_1, B_1, C_1. With the lower level of "A" treatment and no B or C, the symbol is $A_2B_1C_1$; the others follow the same plan. Yates (1937) describes the design.

It should be noted that in fairly simple cases such as the preceding example or those mentioned in Chapter 3, there is considerable latitude in arranging the individual degrees of freedom. When the material grows more complex, as in this example, the arrangement of single-degree comparisons grows more difficult. Usually the comparisons in interactions are more or less "formal;" to get all of them arranged it is necessary to follow rather rigid procedure, and it is difficult to vary the procedure to make specially interesting comparisons. The triple interaction in a $3 \times 3 \times 3$ with its 8 degrees of freedom is a good illustration of this. In making up these degrees of freedom, Yates (1937, Secs. 10–12) sets out the results in 3×3 tables of two treatment combinations for each level of the third. This makes 3 such tables. In our notation the numbers of diseased bulbs are shown in table 52. In arranging degrees of freedom for triple interaction, the diagonals of each division of the table are taken to form the "I" and "J" table (table 53). The "I" diagonals of the A_1 are $B_1C_1 + B_2C_2 + B_3C_3$; $B_1C_2 + B_2C_3 + B_3C_1$; $B_1C_3 + B_2C_1 + B_3C_2$; or 60, 43, and 44. These are the I_1, I_2 and I_3 values. The same procedure is followed with the A_2 and A_3 tables, giving a 3×3 table of I_1 to I_3 vs A_1 to A_3. The "J" diagonals for A_1 are $B_3C_1 + B_2C_2 + B_1C_3$; $B_3C_2 + B_2C_3 + B_1C_1$; $B_3C_3 + B_2C_1 + B_1C_2$. The values are 51, 46, and 50 respectively. Proceeding in the same way with A_2 and A_3, 9 values in all are obtained for AJ combinations.

The six diagonals of each division of table 53 are taken in the same manner. It will be seen that each AI or AJ figure represents 3 original units, and each diagonal of AI and AJ represents 9 units. The 3 diagonal totals from the left of the "I" table give a set which can be used for 2 degrees of freedom of the triple interaction. They are calculated as (Total 1 − Total 2)^2/18 and [Total 1 + Total 2 − 2(Total 3)]^2/54. The three totals of

TABLE 52

	A_1			A_2			A_3		
	B_1	B_2	B_3	B_1	B_2	B_3	B_1	B_2	B_3
C_1	22	17	15	6	10	12	12	12	15
C_2	15	20	11	3	2	6	2	6	4
C_3	16	13	18	7	4	2	7	11	5

TABLE 53

	I				J		
	A_1	A_2	A_3		A_1	A_2	A_3
I_1	60	10	23	J_1	51	21	28
I_2	43	19	28	J_2	46	16	27
I_3	43	23	23	J_3	50	15	19

TABLE 54

Group 1	Group 2	Group 3
$A_1B_1C_1$	$A_1B_1C_3$	$A_1B_1C_2$
$A_1B_2C_3$	$A_1B_2C_2$	$A_1B_2C_1$
$A_1B_3C_2$	$A_1B_3C_1$	$A_1B_3C_3$
$A_2B_1C_2$	$A_2B_1C_1$	$A_2B_1C_3$
$A_2B_2C_1$	$A_2B_2C_3$	$A_2B_2C_2$
$A_2B_3C_3$	$A_2B_3C_2$	$A_2B_3C_1$
$A_3B_1C_3$	$A_3B_1C_2$	$A_3B_1C_1$
$A_3B_2C_2$	$A_3B_2C_1$	$A_3B_2C_3$
$A_3B_3C_1$	$A_3B_3C_3$	$A_3B_3C_2$

the I division diagonals from the right give 2 other pairs. The J division gives 4 others. These are the 8 degrees for triple interaction.

In confounding two of these degrees with blocks, the aim is to make the block comparisons the same as those of the interaction. This is accomplished by taking the 3 sets of 9 from one of the sets of diagonals mentioned and using them in the 3 blocks. Each of these sets will contain all the two-way combinations and will be balanced for A, B, and C, because of the method of making it up.

For the AJ division, the first diagonal from the left is $A_1J_1 + A_2J_2 + A_3J_3$. Tracing the formation back, this combination includes the 9 units; $A_1B_3C_1$, $A_1B_2C_2$, $A_1B_1C_3$, $A_2B_3C_2$, $A_2B_2C_3$, $A_2B_1C_1$, $A_3B_3C_3$, $A_3B_2C_1$, $A_3B_1C_2$. These 9 accordingly are put in the first block. The second diagonal, $A_1J_2 + A_2J_3 + A_3J_1$, give 9 units to be included in the second block; the third, $A_1J_3 + A_2J_1 + A_3J_2$, gives the third block group. These are drawn together and put in logical order in table 54.

The set of 9 first arranged is "group 2" above. Thus the block comparisons are made to coincide with 2 degrees of freedom in triple interaction. Each set may be seen to include every combination of *two* factors, though all three must be taken to get every combination of *three*. The sets were randomized in the plots of each block for the actual field work.

With more than one replication, the confounding might be extended; in a second replication, instead of the left diagonals for AJ, the left (or right) diagonals for AI might be taken; and so on. Thus each degree of freedom for

triple interaction would be confounded in one replication and free in others. It could thus be calculated. However, it is seldom that special interest attaches to particular degrees of freedom for triple interaction.

This gives an idea of the methods necessary in working out confounding schemes. Statistical analysis of results is not difficult. Sums of squares for individual degrees of freedom may be calculated as in the preceding example. The use of diagonal totals of AI and AJ tables has been outlined for triple interaction. Two-way interaction may be calculated from differences of diagonal totals in tables constructed for all combinations of A and B, B and C, A and C. Main effects follow lines of the previous example. Where the treatment is in ascending steps, linear and curvilinear components may be separated.

It is not even necessary to calculate individual degrees unless they are of special interest. If we are once assured that confounding has been properly done, analysis can be carried out as in previous work. Sums of squares for main effects and two-way interactions can be calculated quickly in the ordinary way, using a correction factor. Triple interaction sum of squares may be secured by subtraction from the total. The sum of squares for blocks may be calculated, and deducted from the sum for triple interaction. The remaining 6 degrees of triple interaction gave an estimate of error. If single degrees are calculated for two-way interaction, unimportant constituents could be included in the error.

The summary of analysis is shown in table 55.

30. A triple lattice experiment

Preliminary tests had been made of large numbers of barley strains for resistance to an insect pest. It was desired to use the more promising strains

Table 55

Source of variation	Degrees of freedom	Sum of squares	Mean square
Between treatments	6	—	—
Main effects:			
Effect of A	2	549.56	274.78**
Effect of B	2	2.89	1.44
Effect of C	2	160.89	80.44†
Two-way interactions lumped	12	73.33	6.11
Triple interaction	8	—	—
Blocks	2	8.67	4.34
Remainder, used as error	6	99.33	16.56

** High significance. † Near significance.

in a more precise test with limited replication. A lattice design was suggested as likely to get the most out of the material. Since 3 replications, and probably not more, could be used, the triple lattice fitted the case a little better than would a simple lattice giving 2 or 4 replications. A triple lattice compares more pairs directly than a simple lattice. The design is described by Cochran and Cox (1957) or Cox, Eckhardt, and Cochran (1940).

It is best that the number of varieties or treatments be a perfect square to make up these designs. With a great many treatments, omission of a few or inclusion of a few more is usually not difficult. In this case it was fairly easy to round out the number of strains for further test to make 100. Accordingly 100 were selected and numbered from 1 to 100. The sets are fairly easy to work out, if the n^2 treatment numbers are set down in consecutive order in an $n \times n$ square. The first n letters are then written in to give a Latin square; each letter must occur once and only once in each row and each column. The rows give one group of sets, the columns another, the Latin letters another. This insures that pairs of treatments occurring together in one set will not occur in another. It is not necessary that the Latin square be a random one, such as would be needed for field plots; a systematic rotation of letters can be used. In this design the square is used only for making up the sets; randomization in field blocks comes later. In the present case, however, a random square was written. The square was written to make up sets for strains 1 to 100 (table 56).

The 10 sets of 10 strains for the first replication are made up from the rows: 1, 2, 3 · · · 10; 11, 12 · · · 20; etc. For the second replication they are made up from the columns: 1, 11, 21 · · · 91; 2, 12 · · · 92, etc. For the third replication they are made up from the letters. The first set is made up from the A's: 1, 14, 26, 37, 42, 60, 68, 73, 85, 99. The second set is made up from the B's: 2, 11, 29 · · · 95. The other sets follow the same rule.

Thus 3 groups of 10 sets, each of 10 strains, are made up. They fulfill the requirements for this design, and the appropriate process of analysis

Table 56

1A	2B	3C	4D	5E	6F	7G	8H	9I	10J
11B	12E	13J	14A	15H	16C	17F	18D	19G	20I
21C	22I	23D	24G	25J	26A	27H	28E	29B	30F
31D	32F	33B	34C	35I	36E	37A	38G	39J	40H
41E	42A	43G	44F	45C	46I	47J	48B	49H	50D
51F	52C	53I	54H	55G	56B	57D	58J	59E	60A
61G	62J	63F	64I	65D	66H	67E	68A	69C	70B
71H	72G	73A	74B	75F	76J	77C	78I	79D	80E
81I	82H	83E	84J	85A	86D	87B	88C	89F	90G
91J	92D	93H	94E	95B	96G	97I	98F	99A	100C

can be used with them. The next concern is placing them in the field. Each replication should be in one solid area. This area should be of 100 plots, divided into 10 blocks, each of 10 contiguous plots. There are no other special restrictions on shape of replication areas or blocks. The 10 sets of each replication must be randomized among the 10 blocks of its area, and the 10 strains of each set among the 10 plots of each block.

There are thus 3 replication areas, 30 blocks, and 300 plots. Insect attack was rather epidemic in character early in the season, and percentage of leaves affected was taken as the criterion of infestation. A diagram of the layout, with strain numbers above and percentages of attack below for each plot, is shown in tables 57, 58, and 59.

Analysis proceeds as follows: first the ordinary calculation as for a ran-

TABLE 57. Replication 1

										Block total
8	10	9	7	6	3	5	1	4	2	
82	68	52	73	79	53	70	68	84	80	709
53	51	58	55	54	56	60	59	57	52	
49	82	66	54	28	59	51	43	55	48	535
68	64	70	63	65	67	61	69	66	62	
28	47	41	40	61	30	52	35	46	49	429
81	84	88	85	89	86	82	90	87	83	
66	65	26	93	70	41	27	40	42	44	514
97	91	95	100	94	92	98	93	96	99	
69	48	35	74	23	70	47	45	59	86	556
14	20	11	16	13	17	12	18	19	15	
50	48	60	61	62	47	44	50	61	69	552
71	78	73	75	80	76	74	77	79	72	
39	35	44	56	58	39	35	30	28	42	406
25	28	21	24	26	30	29	22	27	23	
44	86	40	40	38	46	40	48	42	51	475
39	31	36	38	34	33	40	35	37	32	
38	38	56	50	56	41	49	50	47	65	490
48	43	46	42	41	44	50	45	47	49	
42	46	38	52	55	37	36	41	66	70	483
										5149

TABLE 58. Replication 2

										Block total
49	29	99	69	89	79	9	59	39	19	
47	43	81	34	87	24	56	51	36	60	519
70	80	50	100	20	40	30	60	90	10	
31	57	32	64	57	50	46	65	35	68	505
73	63	83	23	13	43	53	33	93	3	
39	43	45	52	59	50	34	37	55	40	454
26	16	56	6	76	86	96	46	36	66	
37	55	78	73	46	59	74	36	60	58	576
58	98	28	78	68	38	88	8	48	18	
61	56	89	37	24	54	20	78	75	57	553
62	52	2	42	22	82	12	32	92	72	
47	45	70	52	50	23	41	49	43	35	455
45	25	95	5	75	15	65	35	85	55	
49	50	40	78	74	56	61	52	66	54	580
44	94	4	34	54	84	14	74	64	24	
40	29	63	60	24	48	38	32	40	43	417
81	1	61	51	31	71	91	41	11	21	
58	75	36	57	48	34	43	55	60	36	502
97	27	87	67	77	57	7	47	37	17	
54	50	43	22	33	60	58	66	47	52	485
										5046

domized block experiment is carried out. Variance for treatment, replication, and interaction are calculated, ignoring the confounding. This gives results as in table 60.

The replication differences are clear-cut, as all varieties are included in each. The variety differences will be adjusted later. The sum of squares for interaction is to be reduced by estimating small block variance. Small block differences are estimated here by taking all varieties in each block and comparing results with the same varieties over the rest of the experiment. For instance, Block 1 has percentages totaling 709. The same varieties, 1 to 10, in various blocks of the second replication, total 659. In the third, these varieties total 583. The estimate of deviation of this block from the general mean is $(2 \times 709) - 659 - 583$, or $+176$. Block 2, in the same way, has a deviation of $+72$. We secure 30 such deviations, total-

TABLE 59. Replication 3

										Block total
37	85	68	14	73	42	1	60	99	26	
40	67	31	43	40	40	72	50	83	45	511
69	52	16	100	3	77	34	21	45	88	
31	40	71	57	38	33	69	39	42	25	445
33	56	2	48	70	87	95	11	74	29	
48	60	68	50	37	47	32	69	43	67	521
10	91	25	84	13	39	62	58	76	47	
70	46	53	48	69	58	50	72	51	82	599
81	20	64	78	22	46	35	53	97	9	
66	50	43	35	45	37	58	44	47	67	492
5	59	83	80	41	36	12	67	28	94	
50	42	46	61	55	54	42	22	90	28	490
40	8	82	66	27	15	49	54	71	93	
65	67	23	47	53	68	45	34	63	51	516
38	90	72	19	43	24	7	61	55	96	
50	33	47	60	43	59	56	28	45	47	468
31	86	23	57	92	4	79	65	50	18	
39	42	39	41	35	52	27	40	22	44	381
17	44	63	98	51	6	75	89	30	32	
48	36	32	38	41	43	47	56	34	46	421
										4844

TABLE 60

	Degrees of freedom	Sum of squares	Mean square
Between replications	2	482	241
Between varieties	99	52,039	526
Interaction, variety × replication	198	12,867	65

ing 0, but having a net negative or positive sum in each replication. These replication sums are +408, +99, −507, respectively. The analysis from the bulletin cited (Cox, Eckhardt, and Cochran 1940) was used; the analysis from Cochran and Cox (1957) is easier, though less detailed in explanations.

We square and sum all the deviations and divide the sum by 2 rk, where r is the number of replications (3) and k is the number of plots in a block. Hence 2 rk is 60. From the quotient is subtracted the sum of squares of the 3 replication net sums of deviations divided by 2 rk^2 or 600. This is expressed as $[(+176)^2 + (+72)^2 + \cdots + (\text{deviation block } 30)^2]/60 - [(\text{Net sum deviations Replication } 1)^2 + (\text{D.R. } 2)^2 + (\text{D.R. } 3)^2]/600$ which comes out 4,583. That is our estimate of the part of the "interaction" sum of squares which is due to block variance.

The analysis may now be restated: subtracting 4,583 from 12,867 to leave 8,284 in *error*, as in table 61. The degrees of freedom among 30 blocks are 27 instead of 29, because 2 belong to replications. Quite a reduction in error estimate has been made possible.

Next we compute weighting coefficients for adjustment of variety means before testing significance of differences. If E is the error variance (48) and B is the block variance (170), $W = 1/E$ or 0.02083; and $W' = 2/(3B - E)$ or 0.00433. The weighting factor (W.F.) is $2(W - W')/(2W + W')$ or 0.03300/0.04599 which comes out 0.71755. This is divided by 2 rk or 60, giving 0.01196. This in turn is multiplied by the 30 block deviations already calculated, for adjustment of varietal means; this is shown below.

Lastly we compute two values for standard error of variety mean differences. One is to test differences between adjusted value of means of varieties occurring once in the same block. The other is for varieties not in the same block. This first is calculated as:

$$\sqrt{2V/rk\left[\frac{6W}{2W + W'} + k - 2\right]}$$

where V is 48, the error variance. This comes out as

$$\sqrt{\left(\frac{96}{30}\right) \times 10.72}$$

or $\sqrt{34.30}$ or 5.86. The second is

Table 61

	Degrees of freedom	Sum of squares	Mean square
Between replications	2	482	241
Between varieties (ignoring blocks)	99	52,039	526
Blocks (adjusted)	27	4,583	170
Intra-block error	171	8,284	48

$$\sqrt{2V/rk\left[\frac{9W}{2W - W'} + k - 3\right]}$$

which comes out $\sqrt{35.46}$ or 5.96. The 30 block deviations and their products by the weighting factor are shown in table 62.

For special tests of significance of differences, variety means should be adjusted by subtracting these weighted deviations for the block in which they fell. For instance, variety 1 occurred in Blocks 1, 19, and 21. The weighted deviations for these blocks are +2.10, −0.74, and −0.71, totaling +0.65. The average percentage for variety 1 is (68 + 75 + 72)/3, or 71.67, with the adjustment this becomes 71.67 − 0.65, or 71.02. Variety 65 occurred in Blocks 3, 17, and 29 with adjustments of +1.21, +1.02, and −3.00, totaling −0.77. With this adjustment its mean of (61 + 61 + 40)/3, or 54.00, becomes 54.00 − (−0.77) or 54.77. The adjusted difference between these two is 71.02 − 54.77, or 16.25. This is significant when compared to 5.96, the standard error of a difference between two varieties not found together in any block. The adjustment is usually small.

The "F" test may be made with the simple analysis shown first, and if it is significant (as is true here), no other test is needed. If it is not significant, a further complex calculation described by Cox, Eckhardt, and Cochran (1940) may be made leading to a more accurate "F" test. In the present case, "F" of the simple test is 526/65, highly significant, and the more complex test gives an "F" not much higher.

Following the methods of Cox, Eckhardt, and Cochran (1940) as to comparison of efficiency, this experiment is estimated as 20 percent more

TABLE 62. Adjusted block deviations

Block number	Deviation	Deviation × weight (0.01196)	Block number	Deviation	Deviation × weight (0.01196)
1	+176	+2.10	16	−51	−0.61
2	+72	+0.86	17	+85	+1.02
3	+101	+1.21	18	−86	−1.03
4	+91	+1.09	19	−62	−0.74
5	+107	+1.28	20	0	0.00
6	+5	+0.06	21	−59	−0.71
7	−46	−0.55	22	−10	−0.12
8	−70	−0.84	23	+58	+0.69
9	−40	−0.48	24	+129	+1.54
10	+12	+0.14	25	+8	+0.10
11	−21	−0.25	26	−56	−0.67
12	+20	+0.24	27	+60	+0.72
13	−17	−0.20	28	−80	−0.96
14	+139	+1.66	29	−251	−3.00
15	+92	+1.10	30	−306	−3.66

efficient than a simple randomized block layout. This means that the 3 replications were almost as good as 4 would have been with the simple plan. A nice gain is thus shown for a few hours paper work and some extra care in handling the test.

31. A lattice square experiment

In studying insecticidal control of boll weevils and aphids on cotton, it was desired to test 16 treatments. This is one of the numbers that can readily be arranged in a lattice square design. The experimenter desired to try it, though with so few treatments it is usually questionable whether a gain will be made by this plan.

First it is necessary to arrange the treatment combinations. In this scheme the sets are made up with certain restrictions, as in the ordinary lattice and triple lattice. Added complexity is given by added restrictions on placing the treatments in plots in the field. With the lattices such as in the previous example, the replication areas and blocks within these areas must constitute definite contiguous plot groups, but with no restriction on their shape. In the lattice square, on the other hand, each replication area contains p^2 plots physically arranged in a square of p rows and p columns. The design can be arranged for 16, 25, 49, 64, 81, or 121 treatments; that is, for $p = 4, 5, 7, 8, 9,$ or 11; not for $K = 6, 10,$ or 12. Where p is even, $p + 1$ replications must be used; every one of the possible pairs of treatments will occur together twice, in some row and some column. Where p is odd, the design may be arranged with only $(p + 1)/2$ replications; each pair will occur once, in some row or some column.

The design is planned by making up an orthogonal square of p^2 units, making several groups of sets. For each replication, rows of plots are made up from one set, columns from another. After arranging these, randomization of rows and columns is carried out in placing in the field. These steps will be illustrated in the problem discussed.

An orthogonal 4 × 4 square was written by taking 3 Latin squares from Fisher and Yates' book of tables (1963). These squares are given in table 63. All were superimposed on one square. For the letters of the first, capitals were used; for those of the second, small letters; for those of the third, A′, B′, etc. The 16 positions were also numbered, corresponding to treatment numbers. The resulting square is shown in table 64. It will be noticed

Table 63

A	B	C	D	A	B	C	D	A	B	C	D
B	A	D	C	C	D	A	B	D	C	B	A
C	D	A	B	D	C	B	A	B	A	D	C
D	C	B	A	B	A	D	C	C	D	A	B

TABLE 64

1	A	a	A′	2	B	b	B′	3	C	c	C′	4	D	d	D′
5	B	c	D′	6	A	d	C′	7	D	a	B′	8	C	b	A′
9	C	d	B′	10	D	c	A′	11	A	b	D′	12	B	a	C′
13	D	b	C′	14	C	a	D′	15	B	d	A′	16	A	c	B′

TABLE 65

1	2	3	4	1	6	11	16	1	8	10	15
5	6	7	8	7	4	13	10	2	7	9	16
9	10	11	12	12	15	2	5	3	6	12	13
13	14	15	16	14	9	8	3	4	5	11	14

1	5	9	13	1	7	12	14
6	2	14	10	8	2	13	11
11	15	3	7	10	16	3	5
16	12	8	4	15	9	6	4

that each set of letters constitutes a Latin square, with each letter occurring once and only once in each row and each column. It is also true that each small letter occurs once and only once with each capital letter, and with each letter of the A′, B′, C′, D′ set. Each set of letters has this arrangement with respect to the other sets.

From this *orthogonal square*, the 16 treatment numbers can be arranged in 5 groups of sets. In each group there are 4 sets of 4. One group is made up from rows, one from columns, one from capital letters, one from small letters, one from A′, B′, etc. For example, with capital letters: A occurs with 1, 6, 11, 16; B with 2, 5, 12, 15, etc. These groups were made up as in Chapter 3, and are written below.

Rows: 1, 2, 3, 4; 5, 6, 7, 8; 9, 10, 11, 12; 13, 14, 15, 16.

Columns: 1, 5, 9, 13; 2, 6, 10, 14; 3, 7, 11, 15; 4, 8, 12, 16.

A's, B's, etc.: 1, 6, 11, 16; 2, 5, 12, 15; 3, 8, 9, 14; 4, 7, 10, 13.

a's, b's, etc.: 1, 7, 12, 14; 2, 8, 11, 13; 3, 5, 10, 16; 4, 6, 9, 15.

A′, B′, etc.: 1, 8, 10, 15; 2, 7, 9, 16; 3, 6, 12, 13; 4, 5, 11, 14.

These sets fulfill the requirements for a lattice square; all possible pairs occur together and none twice. In the 5 replications, each set must occur twice, once as a row and once as a column.

The first replication had rows from the first group of sets, columns from the second group. The second had rows from the third group, columns from the fourth. The other three replications took their rows and columns from the fifth and first groups, the second and third, the fourth and fifth respectively. The rows and columns were as noted in table 65.

In setting these out in the field, rows and their columns were randomized in location. This gives a degree of randomness, yet keeps the sets together. With the first replication a random drawing gave the order 3, 1, 4, 2 to apply to rows. The rows were accordingly placed; on top the third, then

the first, fourth, and second in order (table 66). The columns were then randomized, the drawing giving the order 2, 4, 1, 3 (table 67). In this order, the plots were arranged in the first replication. The other replications were treated similarly. Through misunderstanding, the experimenter did not randomize all replications as completely as the first, but results will serve for illustration.

Records were taken on boll weevil attack, aphid counts, and yield, but analysis will be stated only for aphid counts. The field diagram, with treatment numbers above and aphids per leaf during the most important period below, is as shown in table 68. Treatment totals are: 1, 184.0; 2, 33.9; 3, 73.2; and so on.

Row and column totals for each square (20 of each in all) must likewise be accumulated.

Analysis was made as outlined by Yates (1940); it is described also by Cochran and Cox (1957). The principle is (as with the ordinary lattice) the adjustment of area values allowing for treatments in each area, followed by the adjustment of treatments for areas. The row and column totals were computed for each replication. Replication and treatment totals are secured.

A table of parallel columns is arranged, with a line for each of the 16 treatments. The first column is of treatment designations. The second is headed T and contains the total value for the given treatment. The third is R, and contains the total for the rows containing each treatment. For example, the row totals for rows containing treatment 1 had total values of 90.3, 87.2, 98.3, 119.3, 110.6; the 5 total 505.7. The fourth column is C and contains the total for columns associated with the given treatment. For treatment 1, this is 501.2. The fifth column, D_o, is R − C; for No. 1, this is +4.5. The sixth column, L'_o, is $(p \times T) - [(R \times (p + 1)] +$ (total of all plots). For treatment 1, this is $(4 \times 184.0) - (505.7 \times 5) +$ 1814.7, or 22.2. The next column, J_o, is $L'_o + D_o$, or +26.7. The next, K_o, is $J_o + (p - 1)D_o$, or +40.2. The eighth, M_o, is $K_o + D_o$, or 44.7. The next, M'_o, should check M_o; it is $(p \times T) - [(p + 1) \times C] +$ (total of all plots), or for treatment 1, +44.7. The last column is left for adjusted treatment values, to be calculated later.

The total sum of squares of deviation, the sum for replications, the sum for treatments ignoring rows and columns, and the remainder sum are calculated as in simpler designs. The sum of squares for rows, eliminating

Table 66

9	10	11	12
1	2	3	4
13	14	15	16
5	6	7	8

Table 67

10	12	9	11
2	4	1	3
14	16	13	15
6	8	5	7

TABLE 68

				10 51.0	12 63.9	9 53.5	11 54.1	5 10.3	12 45.2	15 10.0	2 7.5
				2 7.8	4 17.5	1 45.1	3 19.9	10 43.6	7 10.2	4 8.3	13 24.2
				14 41.0	16 55.6	13 33.5	15 8.2	16 23.0	1 36.9	6 4.3	11 23.0
				6 4.5	8 10.1	5 5.8	7 8.1	3 17.0	14 24.4	9 33.8	8 10.9
10 39.9	15 9.4	8 15.3	1 33.7	16 36.1	12 31.2	8 11.7	4 11.9	3 12.6	16 28.1	5 9.3	10 21.5
9 56.2	16 30.8	7 7.2	2 6.1	11 32.5	15 5.1	3 12.2	7 5.9	6 3.3	9 31.3	4 16.3	15 5.0
12 28.9	13 26.8	6 6.8	3 11.5	1 36.6	5 5.7	9 38.1	13 38.9	12 40.0	7 8.9	14 30.0	1 31.7
11 33.5	14 43.0	5 7.5	4 12.6	6 5.6	2 5.2	14 23.0	10 21.0	13 27.1	2 7.3	11 26.5	8 13.7

treatments, is given by the sum of squares of the L'_o column, with divisor $(p)^3 \times (p + 1)$. The M_o column in the same way gives the sum of squares for columns, eliminating treatments. The sum of squares of the J_o column, with divisor $(p)^3 \times (p - 1)$, gives the sum of squares for rows, eliminating varieties and columns. The K_o column in the same manner gives the sum for columns, eliminating both treatments and rows.

The sum for rows eliminating columns and treatments, plus that for columns eliminating treatments only; or columns eliminating rows and treatments, plus that for rows eliminating treatments only; should be equal. Either one can be subtracted from the "remainder" sum previously calculated, to give the sum of squares for error.

The analysis so far is summarized in table 69.

Adjustments for treatment totals are then developed. Weights used are w_i reciprocal of error mean square, $1/36.34 = 0.0275$, and w_r, which is $(p - 1)/[(p \times \text{adjusted row mean square}) - (\text{error mean square})]$ and comes out 0.0085. Another weight, w_c, could be computed using column mean square, but this is nonsignificant and is regarded as nonexistent.

The weight "lambda" (λ) is $(w_i - w_r)/[w_r + w_c + (p - 1)w_i]$ or 0.209; $\lambda/p = 0.052$. The adjustment for rows in any treatment is $L'_o \times \lambda/p$.

TABLE 69

	Degrees of freedom	Sum of squares	Mean square
Total	79	19,229.09	—
Replications	4	1,180.48	295.12
Treatments (ignoring confounding)	15	15,188.37	1,012.56
Remainder	60	2,860.24	—
Columns, eliminating treatments (M_o)	15	322.82	—
Rows, eliminating C and T(J_o)	15	1,447.14	96.48
Rows, eliminating T(L'_o)	15	1,381.86	—
Columns, eliminating C and R(K_o)	15	388.10	25.87
Error	30	1,090.28	36.34

For treatment 1 this is $0.052 \times (+22.2)$ or about $+1.2$; the total is accordingly adjusted from 184.0 to 185.2. The adjustment for treatment 2 is negative. Some other treatments have larger adjustments; the largest is to No. 6, which rises from 24.5 to 45.9. This treatment occurred in rows of low infestation, and its rating is thus raised.

A second adjustment, "mu" (μ), may be calculated for column tendencies using M_o, but here column variation is regarded as nonexistent.

The standard error of a treatment total is estimated as

$$\sqrt{(1 + \lambda + \mu) \times \text{No. replications} \times \text{error mean square}},$$

here

$$\sqrt{(1 + 0.209 + 0) \times 5 \times 36.34} \text{ or } 14.8.$$

Several other notes may be made. A simple randomized block analysis would give a standard error of a mean difference as 4.4. Squaring these two standard errors for comparison, we find the lattice square to be rated as about 110% as efficient as the randomized block would have been. Use of the square gave a gain, but not as much gain as another replication would have done with the simpler plan. The variation may be seen to be directional, rows showing differences and columns no particular difference. This seems to occur often with insect populations. The boll weevil infestation showed smaller but significant treatment differences; both rows and columns varied, the rows more; the lattice square gave about 50% gain.

If there are only $(p + 1)/2$ replications in this design, which is possible if p is odd, analysis is a little simpler. Rows and columns can be adjusted only for treatments, not for each other. The adjustment is carried out much as with the ordinary lattice, and further procedure is outlined by Yates. The present method of analysis represents a definite gain in efficiency over the earlier method proposed by Yates.

32. Discriminant function and diagnostic characters

The problem of diagnostic characters has been touched on in Chapter 1. Two groups of beetles which were similar but were believed to be distinct were studied. Four measurements were made on each; prothorax width and length, elytron width and length. Often when a single measurement will not serve to discriminate, a combination of several can be found which will do so. Fisher (1954) outlines the calculation.

In the present case this did not prove true. The individual measurements showed overlapping, and the "discriminant functions" calculated from several measurements also overlapped. It was noted that the measurements were *correlated;* that if prothorax length was high or low in an individual beetle, elytron length was also high or low as a rule. This suggested the idea that the *ratios* of two measurements might be quite stable. These ratios were more promising than individual measurements. While no single ratio separated the groups, a combination was found by least square methods which did separate them. The procedure will serve to illustrate calculation of the discriminant function.

The six possible ratios were calculated and tabled for 10 specimens of each group. The ratios were taken with the larger measurement in the numerator so that each ratio was greater than 1. For example, in group A, specimen No. 1, prothorax width (PW) was 11.50 units, prothorax length (PL) was 9.50. The ratio PW/PL was 11.50/9.50 or 1.21. EL denotes elytron length.

Measurements selected as most promising of the 6 were PW/PL and EL/PW. They are shown in table 70.

TABLE 70

	Group A		Group B	
	PW/PL(X_1):	EL/PW(X_2)	PW/PL(X_1):	EL/PW(X_2)
	1.21	2.04	1.24	2.08
	1.25	2.16	1.30	2.00
	1.21	2.13	1.25	2.02
	1.18	2.06	1.25	2.04
	1.18	2.17	1.26	2.00
	1.23	2.13	1.24	2.08
	1.22	2.21	1.27	2.02
	1.22	2.18	1.27	2.04
	1.19	2.21	1.25	2.13
	1.20	2.27	1.26	2.00
Sum	12.09	21.56	12.59	20.41
Mean	1.209	2.156	1.259	2.041

The sum of squares of deviations of PW/PL and EL/PW (termed X_1 and X_2 respectively for convenience) is calculated for the series of 20 without division as in simple correlation. The sum of products of deviations of X_1 and X_2 is also calculated. The difference in the average X_1 for group A and group B is also calculated; it is 1.209 − 1.259, or −0.050. This is called d_1. The difference, in the same way, in X_2, is 2.156 − 2.041, or +0.115 (d_2). Note that the difference must be taken the same way both times; we must say A − B or B − A and stick to it. Either "*d*" may be positive or negative. This time we have one of each.

The equations are set up as for multiple regression (Chapter 1):

$$S(x_1^2) + S(x_1x_2) = d_1 \qquad S(x_1x_2) + S(x_2^2) = d_2$$

In this case $S(x_1^2)$ is

$$(1.21)^2 + (1.25)^2 + \cdots + (1.26)^2 - [(12.09 + 12.59)^2/20]$$

which is 0.0199. $S(x_1x_2)$ is

$$(1.21 \times 2.04) + \cdots + (1.26 \times 2.00) - [(12.09 + 12.59) \times (21.56 + 20.4)]/20,$$

or − 0.0322. $S(x_2^2)$ is 0.1267. The equations for the coefficients of X_1 and X_2 are:

$$0.0199b_1 - 0.0322b_2 = -0.050$$

$$-0.0322b_1 + 0.1267b_2 = +0.115.$$

Solving, b_1 comes out as −1.775; b_2 as +0.456.

Next the coefficients are used on the ratios X_1 and X_2 to calculate the discriminant function for each beetle. For the first, the function is (−1.779 × 1.21, or −2.15) + (+0.456 × 2.04, or +0.93), which is −1.22. The others are calculated in order, with the results for the 10 specimens of each group as in table 71. If d_1 and d_2 had been taken as B − A instead of A − B, we would have had positive functions with a little clearer picture. However, it is easy to see that all of group B have values of −1.25 or larger negatives, while all of group A have −1.24 or smaller negative values. The discrimination is clear-cut, while either ratio alone showed overlapping. This process may easily be extended to three or more measurements or ratios, in the manner typical of multiple regression calculations.

33. Multiple means

In data closely similar to some studied by the author there are 5 treatments in 4 replications with 12 degrees of freedom for error (randomized

TABLE 71

Group A	Group B
−1.22	−1.25
−1.24	−1.40
−1.18	−1.30
−1.16	−1.29
−1.11	−1.33
−1.21	−1.25
−1.16	−1.33
−1.17	−1.32
−1.10	−1.25
−1.09	−1.33

blocks). The error variance is 10.24, the standard error of a treatment mean ($s_{\bar{x}}$) is $\sqrt{10.24/4}$ or 1.60. The standard error of a mean difference is $\sqrt{2} \times 1.6$ or 2.26. Since the "t" for 5% with 12 d.f. is 2.18, the "least significant difference" is 2.18 × 2.26 or 4.93. (This result might also be secured by taking 3.08 × $s_{\bar{x}}$.)

The treatment yield means are ranked in order as follows:

A 28.4[a]	C 26.1[ab]	check 21.7[b]
B 28.3[a]	D 24.1[ab]	

Using the "least significant difference," it is apparent that A and B differ from the check, but not from C and D; also, C and D do not differ from the check.

The situation mentioned in Chapter 1, that a difference may appear significant when it is not, because of the number of differences involved, must be considered here. Snedecor (1956, Chapter 1, sec. 10.6) gives Tukey's method for dealing with it. From his table, the multiplier for $s_{\bar{x}}$, with 5 means and 12 d.f. for error, is 4.51 instead of 3.08. This gives us a "D" of (4.51 × 1.6) or 7.2, which is an adequate safeguard against false recognition of differences. In this case it shows *no* difference to be significant.

A more lenient method is given by Duncan (1955). The multiplier is varied with the number of steps apart in the range of means. For 2, 3, 4, and 5 means the multipliers (for 12 d.f.) are 3.08, 3.23, 3.33, 3.36 respectively. Note that the 3.08 is the same as in the simple "least significant difference." For A vs. check, the significant difference is 3.36 × 1.6 or 5.4; the two differ significantly by this standard. For B vs. check, we use 3.33 × 1.6; and so on. The verdict here is that A and B both differ from the check, as in the simple least significant difference.

The means are then given superscripts; two means having the same letter are not significantly different. These methods serve to pick out significant differences where the F-test has shown general significance. In no case

should a significant difference between two means be claimed, where F has not shown significance.

34. Literature cited

Bliss, C. I.
1935a. The calculation of the dosage mortality curve. Ann. Appl. Biol. 22: 134–167.
———
1935b. The comparison of dosage–mortality data. Ann. Appl. Biol. 22: 307–333.
———
1939. Toxicity of poisons applied jointly. Ann. Appl. Biol. 26: 585–615.
Cassil, C. C., Wadley, F. M., and Dean, F. P.
1943. Sampling studies on orchard spray residues in the Pacific Northwest. Jour. Econ. Entomol. 36: 227–231.
Chitwood, B. G., and Blanton, F. E.
1941. Treatment of narcissus bulbs. Jour. Wash. Acad. Sci. 31: 296–308.
Cochran, W. G., and Cox, Gertrude M.
1957. Experimental designs. 12th ed., 611 pp. Wiley, New Jersey.
Cox, Gertrude M., Eckhardt, R. C., and Cochran, W. G.
1940. The analysis of lattice and triple lattice designs in corn varietal tests. Iowa Expt. Sta. Bull. 281.
Duncan, D. B.
1955. Multiple range and multiple F tests. Biometrics 11: 1–42.
Finney, D. J.
1952. Probit analysis. 2nd ed., 318 pp. Cambridge, New York.
Fisher, R. A.
1954. Statistical methods for research workers. 12th ed., 356 pp. Oliver and Boyd, London.
——— and Yates, Frank.
1963. Statistical tables for biological, agricultural, and medical research. 6th ed., 146 pp. Hafner, New York.
Ladell, W. R. S.
1938. Field experiments in the control of wireworms. Ann. Appl. Biol. 25: 341–389.
Martin, J. T.
1942. The problem of evaluation of rotenone-containing plants. Ann. Appl. Biol. 29: 69–81.
Snedecor, G. W.
1956. Statistical methods. 5th ed., 534 pp. Iowa State University Press, Iowa.
Wadley, F. M.
1945. The evidence required to show synergistic action of insecticides, and a short cut in analysis. U.S. Dept. Agr., ET circular 223.
Yates, Frank.
1937. The design and analysis of factorial experiments. Imper. Bur. Soil Sci., Tech. Commun. 35.
———
1940. Lattice squares. Jour. Ag. Sci. 30: 672–687.

INDEX